(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
4 January 2001 (04.01.2001)

PCT

(10) International Publication Number
WO 01/00858 A1

(51) International Patent Classification[7]: C12N 15/82, 5/10, C12Q 1/68, A01H 5/00

(21) International Application Number: PCT/US00/17535

(22) International Filing Date: 26 June 2000 (26.06.2000)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
09/343,775 30 June 1999 (30.06.1999) US

(71) Applicant: WISCONSIN ALUMNI RESEARCH FOUNDATION [US/US]; P.O. Box 7365, Madison, WI 53707-7365 (US).

(72) Inventors: JIANG, Jiming; 5734 Schumann Drive, Madison, WI 53711 (US). DONG, Fenggao; 17 Laramie Court, Madison, WI 53706 (US).

(74) Agent: MUELLER, Lisa, V.; Rockey, Milnamow & Katz, Ltd., Suite 4700, Two Prudential Plaza, 180 North Stetson Avenue, Chicago, IL 60601 (US).

(81) Designated States *(national)*: AE, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, CA, CH, CN, CR, CU, CZ, DE, DK, DM, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, TZ, UA, UG, UZ, VN, YU, ZA, ZW.

(84) Designated States *(regional)*: ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZW), Eurasian

*[Continued on next page]*

(54) Title: DNA SEQUENCES SPECIFIC TO RICE CENTROMERES

```
        10         20         30         40         50
GATCTTTGGA TTGGAAACAG TTAAAGAACA ATATGTGCAT GATGATGATT
        60         70         80         90        100
TTAAAGATGT GTTTTTGCAT TGTAAGGATG GGAAGGCATG GAATAAATTT
       110        120        130        140        150
GTTGTAAATG ATGGTTTTGT GTTTAGAGCT AATAAGCTAT GCATTCCAGC
       160        170        180        190        200
TAGCTCTGTT CGTTTGTTGT TGCTACAGGA AGCACATGGA GGTGGTTTGA
       210        220        230        240        250
TGGGACATTT TGGGGCAAAG AAGACGGAGG ACATACTGGC TGCTCATTTC
       260        270        280        290        300
TTTTGGCCAA AGATGAGGAG AGATGTGGAG AGATTTATTG CTCGCTGCAC
       310        320        330        340        350
GACATGTCAA AAGGCCAAGT CACGCTTAAA TCCACACGAT TTGAAGCCAT
       360        370        380        390        400
ATTTGGGTGA GGGAGATGAG CTTGAGTCGG GGACGACTCA AATGCAAGAA
       410        420        430        440        450
GGGGAGGATG ATGAGGACAT CAGCACCATC TATACATCCA CACCTACACC
       460        470        480        490        500
CACACCATCG CCAACACCAC TTGGCCCTCT TACTCGTGCC AGTGCCCGTC
       510        520        530        540        550
AACTGAACCA TCAAGTAAGT TTATTCTTAA ACTCTTGTCC ATCATATTTA
       560        570        580        590        600
GACAATGGAG ACACGTGCAC TCTTGTTTTG CTTAGGAATG ATGGAGAGGA
       610        620        630        640        650
CCAGAAGCAT AGGGGATTGG TGTAGGCTGG ATTTGGACAG CAAGACAGCA
       660        670        680        690        700
CCAACTTACA ACAACCGCCA TGACTTCATA CAGAGTCCAT TTTAAGCATG
       710        720        730        740        750
CAAGCACTTG ATGGAAAACT CGTCAAGTAT ATTTTTAGAT GGATC
```

*[Continued on next page]*

WO 01/00858 A1

patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).

**Published:**
— *With international search report.*
— *Before the expiration of the time limit for amending the claims and to be republished in the event of receipt of amendments.*

*For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.*

(57) **Abstract:** The present invention relates to nucleic acids which encode a functional centromere from *Oryza sativa*. The nucleic acids of the present invention can be also used to create a plant artificial chromosome.

# DNA SEQUENCES SPECIFIC TO RICE CENTROMERES

This invention was made with United States government support awarded by the following agencies: USDA HATCH 3935. The United States has certain rights in this invention.

## Field of the Invention

The present invention relates generally to molecular biology. In particular, the present invention relates to nucleic acid sequences which encode a centromere from rice.

## Background of the Invention

Among the most distinguishing and characteristic landmarks of chromosomes of higher eukaryotes is the location of the centromere. The centromeric region is the site for mitoitic and meiotic spindle fiber attachment and is responsible for sister chromatid association. Jiang, J., et al., *Proc. Natl. Acad. Sci. USA*, 95: 8135-8140 (1998). Therefore, centromeres play a central role in the process of chromosomal segregation and transmission in cell divisions. *Id.* The molecular organization of centromeres has been studied extensively in yeast, *Drosophila melanogaster*, and humans.

Centromeric regions usually consist of heterochromation and are thought to be highly methylated. Miller, J.T., et al., *Theor. Appl. Genet.*, 96:832-839 (1998). In addition, centromeres show varying amounts of nontranscribed repetitive sequences, which are referred to as satellite DNAs. Haaf, T., et al., *Cell*, 70:681-695 (1992). The predominant class of centromeric DNA, is the alpha-satellite DNA, which is found in diverged form in all centromeres. *Id.* To the extent that it is known, alpha-satellite arrays appear to be uninterrupted by other (nonsatellite) DNA sequences.

In centromeres, naturally occurring satellite arrays range in size from several hundred kbs to several megabases in length. Recent studies, however, suggest that as little as 140 kb of alpha satellite DNA may be sufficient to confer centromere function in human cells. Harrington, John.

J., et al., *Nature Genetics*, 15:345-355 (1997). Unfortunately, satellite DNA of this size has proven difficult to clone and propagate stably in microorganisms using conventional cloning vectors. *Id.* In large part, the difficulty in propagating satellite DNA stems from the tendency of tandemly repetitive DNA to recombine into smaller arrays and this effect increases with the size of the repetitive array. *Id.*

As briefly mentioned hereinbefore, functional centromeric sequences have been isolated and purified from *S. cerevisiae* (see Clark et al., *Nature*, 287:504-509 (1980) and Stinchcomb et al., *J. Molec. Biol.* 158:157-179 (1982)). Episomes carrying the yeast centromeric sequences display proper segregation into daughter yeast cells during mitosis and meiosis, in contrast to autonomous replication sequences plasmids lacking a centromere.

The best characterized centromeric DNAs originated from the budding yeast *S. cerevisiae* (See Clarke and Carbon, *Ann. Rev. Genet.* 19:29-56 (1985)). The DNA region required for centromere function in *S. cerevisiae* is approximately 120 base pair (hereinafter "bp") long and is composed of three conserved domains: CDEI, an 8 bp element (A/G)TCAC(A/G)TG), CDEII, an extremely (about90%) AT-rich region of approximately 80 bp, and CDEIII, a 25 bp element (TGTTT(A/T)TGNTTTCCGAAANNNNAAA). The molecular structure of centromeric DNAs from the fission yeast *Schizosaccharomyces pombe* have also been characterized. Several classes of *S. pombe* moderately repeated DNA elements have been identified which are found only in the centromere regions. These centromere-specific repetitive elements have been designated dg (3.8 kb), dh (4 kb), and yn by Yanagida and co-workers (Nakaseko et al., Embo. J. 5.:1011-1021 (1986); Nakaseko et al., *Nuc. Acid Res.* 15:4705-4715 (1987)), and K (6.4 kb), L (6 kb), and B (1 kb) by Carbon and his colleagues (Clarke et al., *PNAS* 83:8253-8257 (1986); Fishel et al., *Mol. Cell Biol.* 8:754-763 (1988)). The dg element has an AT-rich region and a 600 bp domain containing numerous small direct repeat motifs. Similarly, the dh element has an overall AT content approaching 70% and contains many short direct repeats. No nucleotide similarities to the *S. cerevisiae* CDEs have been found in the *S. pombe* elements.

Attempts to demonstrate that the *S. pombe* centromere-specific repetitive elements can function individually as centromeres have been unsuccessful. However, large restriction fragments (65 to 150 kb) carrying the entire fission yeast centromere regions of chromosome 1 or 3 function as centromeres when introduced into acentric episomes (Hahnenberger et al., *PNAS USA* 86:577-581 (1989)). These results indicate that either fission yeast centromeres are large composite structures that cannot be subdivided, or the functional fission yeast centromere element has not yet been identified.

In contrast to the detailed studies done in *S. cerevisiae* and *S. pombe*, in most eukaryotes, only limited information is available regarding the organization of the centromeres. For example, limited information is known about plant centromeres. Peacock et al., *Proc. Natl. Acad. Sci. USA*, 78:4490-4494 (1981) report the first isolation of a repetitive DNA element from maize knobs. This repetitive DNA element acts as neocentromeres in certain genetic backgrounds. A repetitive DNA element has also been cloned from the centromeres of the supernumerary B chromosomes of maize (see Alfenito, M.R., et al., *Genetics* 135:589-597 (1993) and Kaszas, E., et al., *EMBO J.*, 15:5246-5255 (1996)). Part of this B-specific DNA element shows strong homology to the maize sequences. A 180-bp tandem repeat (pAL1 family) is the major component of the centromeric region of *Arabidopsis thaliana* chromosomes. The genomic organization of this repeat family shares similarities to the alpha satellite DNA at the human centromeres (see Martinez-Sapater, J., et al., *Mol. Gen. Genet.*, 204:417-423 (1986); Simoens, C. R., *Nucleic Acids Res.*, 16:6753-6766 (1988); Maluszynska, J., et al., *Plant J.*, 1:159-166 (1991); Round, E.K., et al., *Genome Res.*, 7:1045-1053 (1997)).

As discussed above, very few putative functional centromeres have been cloned from plants. The cloning of a putative functional centromere from a plant is a necessary first step in the production of artificial chromosomes suitable for use in plants. Artificial chromosomes are man-made linear or circular DNA molecules constructed from essential cis-acting DNA sequence elements that are responsible for the proper replication and partitioning of natural chromosomes (see. Murray et al., *Nature*, 305:189-193 (1983)). The essential elemcnts of an artificial

chromosome are: Autonomous Replication Sequences (ARS) (have properties of replication origins, which are the sites for initiation of DNA replication), (2) centromeres (site of kinetochore assembly and responsible for proper distribution of replicated chromosomes at mitosis and meiosis), and (3) telomeres (specialized structures at the ends of linear chromosomes that function to stabilize the ends and facilitate the complete replication of the extreme termini of the DNA molecule). The use of artificial chromosomes as an alternative to commonly used method of introducing new genetic information into cells is steadily increasing.

Summary of the Invention

The present invention relates to isolated and purified nucleic acids having the nucleotide sequences shown in: SEQ ID NO:1, SEQ ID NO:2, SEQ ID NO:3, SEQ ID NO:4, SEQ ID NO:5 SEQ ID NO:6 and SEQ ID NO:7.

The present invention also relates to a recombinant DNA construct which contains a rice centromere. The centromere contains a number of highly repetitive regions of DNA that have the nucleotide sequence of SEQ ID NO:1, SEQ ID NO:2, SEQ ID NO:3, SEQ ID NO:4, SEQ ID NO:5, SEQ ID NO:6 and SEQ ID NO:7 or combinations thereof. The recombinant DNA construct may also contain a yeast autonomous replication sequence, an autonomous replication sequence from a higher eukaryotic organism, a yeast telomere sequence or a telomere sequence from a higher eukaryotic organism and a selectable marker gene.

The present invention also relates to a plasmid containing the hereinbefore described DNA construct. This plasmid may contain an origin of replication and a selectable marker which functions in bacteria (such as *E. coli*) or in yeast (such as *S. cerevisiae*).

The present invention relates to a plant artificial chromosome vector. The plant artificial chromosome vector of the present invention contains an autonomous replication sequence, two telomere sequences, a centromere sequence having the nucleotide sequence of SEQ ID NO:1,

SEQ ID NO:2, SEQ ID NO:3, SEQ ID NO:4, SEQ ID NO:5, SEQ ID NO:6 and SEQ ID NO:7 or a combination thereof, and at least one selectable marker sequence. The autonomous replication sequence may be from yeast or from a higher eukaryotic organism and the telomere sequence may from yeast or from a higher eukaryotic organism, such as, but not limited to, *Arabidopsis thaliana.*

The present invention also relates to a plant cell transformed with the plant artificial chromosome vector hereinbefore described and to transgenic plants containing said plant cell. The plant cell and plant may be from *Oryza sativa.*

Finally, the present invention relates to a method of identifying centromeric DNA in a higher eukaryotic organism. The method involves hybridizing an isolated nucleic acid selected from the group consisting of SEQ ID NO:1, SEQ ID NO:2, SEQ ID NO:3, SEQ ID NO:4, SEQ ID NO:5, SEQ ID NO:6 and SEQ ID NO:7 and combinations thereof with a sample of DNA from a higher eukaryotic organism and then identifying and isolating the centromeric DNA from said sample.

Brief Description of the Drawings

FIG. 1 shows the nucleotide sequence of pSau3A9.

FIG. 2A - FIG. 2N shows the FISH analysis of rice centromeric DNA elements. The probes were biotinylated and hybridized *in situ* to rice chromosomes or DNA fibers. The probes were detected by fluorescein isothiocyanate-conjugated antibodies (green color) and the chromosomes were stained with propidium iodide (red color). Probes pRCS1 hybridized exclusively to the centromeric regions of the chromosomes from rice (FIG. 2A), rye (FIG.2B), barley (FIG. 2C), sorghum (FIG. 2D), and maize (FIG. 2E). FISH signals also were detected in the centromeric regions of the acrocentric B chromosomes (see arrows) from rye (FIG. 2B) and maize (2E). Similarly, rice centromeric DNA families RCH2 (FIG. 2F), RCH1 (FIG. 2G), RCH3

(FIG. 2H), RCE1 (FIG. 2I), RCE2 (FIG. 2J), and RCS2 (FIG. 2K) all were located in the centromere of every rice chromosome. Two pairs of chromosomes with the strongest signals are indicated by arrows and the third pair with the weakest signals by arrowheads (FIG. 2K). The same metaphase cell (FIG. 2K) was washed under medium (FIG. 2L) and high strigencies (FIG. 2M), and most signals were still discernible (FIG. 2N). The marked array between two arrows is 51 $\mu$m long and represents approximately 151-kb DNA. All bars are 10 $\mu$m.

FIG. 3 shows a Southern blot of the genomic organization of the RCS1 family. Rice genomic DNA was digested with *Sau3*AI (lane 1), *Dpn*II (lane 2), *Hae*III (lane 3), *Msp*I (lane 4), *Hpa*II (lane 5), *Sal*I (lane 6), *Bam*HI (lane 7), *Dra*I (lane 8), *Eco*RI (lane 9), and *Hind*III (lane 10) and probed with pRCS1.

FIG 4 shows the nucleotide sequence of pRCS2. The 639-bp insert of pRCS2 contains four copies of a tandemly arranged repeat. The four members (A-D) range from 155 to 165 bp and share 84-91% sequence identify with one another. F represents the consensus sequence of the four members.

FIG. 5 shows a Southern blot of the genomic organization of the RCS2 family. Rice genomic DNA was digested with *Dpn*II (lane 1), *Sau3*AI (lane 2), *Msp*I (lane 3), *Hpa*II (lane 4), *Sal*I (lane 5), and *Hae*III (lane 6), and probed with pRCS2.

FIG. 6A - FIG. 6B show a Southern blot of the conservation of the RCH1 and RCE1 families in *Gramineae* species. Genomic DNA from sorghum (lane 1), maize (lane 2), sugar cane (lane 3), *Ag. intermedium* (lane 4), barley (lane 5), oats (lane 6), rye (lane 7), wheat (lane 8), *Ae. Squarrosa* (lane 9), rice (lane 10), bamboo (lane 11), *Pharus sp.* (lane 12), *J. effusus* (lane 13), *C. alternifolius* (lane 14) and *A. thaliana* (lane 15) was digested with *Hind*III and probed with pRCH1 (FIG. 6A) and pRCE1 (FIG. 6B).

FIG. 7 shows the nucleotide sequence of RCS1.

FIG. 8 shows the nucleotide sequence of RCS2.

FIG. 9 shows the nucleotide sequence of RCH1.

FIG. 10 shows the nucleotide sequence of RCH2.

FIG. 11 shows the nucleotide sequence of RCH3.

FIG. 12 shows the nucleotide sequence of RCE1.

FIG. 13 shows the nucleotide sequence of RCE2.

Detailed Description of the Invention

Background

The present invention relates to cloned centromeric DNA from *Oryza sativa* (rice). More specifically, the inventors of the present invention have discovered that the cloned centromeric DNA of the present invention contains seven (7) different repetitive regions of complex DNA. These seven (7) repetitive regions are referred to herein as follows: RCS1, RCS2, RCH1, RCH2, RCH3, RCE1 and RCE2.

The present invention relates to isolated and purified nucleic acids for each of the seven (7) different repetitive regions of centromeric DNA from *Oryza sativa*. The nucleic acids of the present invention encode a functional centromere from *Oryza sativa*.

The present invention further relates to the use of the nucleic acids of the present invention as primers and probes to identify centromeric DNA from other plants and animals. In addition, the nucleic acid sequences disclosed herein can be used to create a plant artificial chromosome vector.

Definitions

Units, prefixes, and symbols can be denoted in the SI accepted form. Numeric ranges are inclusive of the numbers defining the range. Unless otherwise indicated, nucleic acids are written left to right in 5' to 3' orientation, respectively. The headings provided herein are not limitations of the various aspects or embodiments of the invention which can be had by reference to the specification as a whole. Accordingly, the terms defined immediately below are more fully defined by reference to the specification as a whole.

As used herein, the term "plant" includes reference to whole plants, plant organs (e.g., leaves, stems, roots, etc.), seeds and plant cells and progeny thereof. The class of plants which can be used in the methods of the present invention are generally as broad as the class of higher plants amenable to transformation techniques, including both monocotyledonous and dicotyledonous plants.

As used herein, the term "transformation" or "transfection" means the acquisition in cells of new DNA sequences through incorporation of added DNA. This is the process by which naked DNA, DNA coated with protein, or whole artificial chromosomes are introduced into a cell, resulting in a heritable change.

As used herein, the term "host" means any organism that is the recipient of a replicable plasmid or vector comprising a plant artificial chromosome. Preferably, host strains used for cloning are free of any restriction enzyme activity that might degrade the foreign DNA used. Preferred examples of host cells for cloning which are useful in the present invention are bacteria, such as *Escherichia coli*, *Bacillus subtilis*, *Pseudomonas*, *Streptomyces*, *Salmonella*, and yeast cells such as *S. cerevisiae*. Host cells which can be targeted for expression of a plant artificial chromosome may be plant cells of any source, such as, but not limited to, *Arabidopsis*, maize, rice, sugarcane, sorghum, barley, soybeans, tobacco, wheat, tomato, potato or citrus.

As used herein, the term "linker" means a DNA molecule, generally up to 50 or 60

nucleotides long and synthesized chemically, or cloned from other vectors.

As used herein, the term "plasmid" or "vector" (such as a cloning vector or expression vector)" refers to a closed covalently circular extrachromosomal DNA or linear DNA which is able to autonomously replicate in a host cell and which is normally nonessential to the survival of the cell. A wide variety of plasmids and other vectors are well known and commonly used in the art.

As used herein, "heterologous" when used to describe nucleic acids or polypeptides refers to nucleic acids or polypeptides that originate from a foreign species, or, if from the same species, are substantially modified from their original form. For example, a promoter operably linked to a heterologous structural gene is from a species different from that from which the structural gene was derived, or, if from the same species, one or both are substantially modified from their original form.

As used herein, "isolated" includes reference to material which is substantially or essentially free from components which normally accompany or interact with it as found in its naturally occurring environment. The isolated material optionally comprises material not found with the material in its natural environment.

As used herein, "nucleic acid" includes reference to a deoxyribonucleotide or ribonucleotide polymer in either single- or double-stranded form, and unless otherwise limited, encompasses known analogues of natural nucleotides that hybridize to nucleic acids in a manner similar to naturally occurring nucleotides. Unless otherwise indicated, a particular nucleic acid sequence includes the complementary sequence thereof.

As used herein, "operably linked" includes reference to a functional linkage between a promoter and a second sequence, wherein the promoter sequence initiates and mediates transcription of the DNA sequence corresponding to the second sequence. Generally, operably

linked means that the nucleic acid sequences being linked are contiguous and, where necessary to joint two protein coding regions, contiguous and in the same reading frame.

As used herein "recombinant" includes reference to a cell, or nucleic acid, or vector, that has been modified by the introduction of a heterologous nucleic acid or the alteration of a native nucleic acid to a form not native to that cell, or that the cell is derived from a cell so modified. For example, recombinant cells express genes that are not found within the native (non-recombinant) form of the cell or express native genes that are otherwise abnormally expressed, under expressed or not expressed at all.

As used herein, a "recombinant DNA construct" is a nucleic acid construct, generated recombinantly or synthetically, with a series of specified nucleic acid elements which permit transcription and translation of a particular nucleic acid in a target cell. The DNA construct can be part of a plasmid, vector, virus, or nucleic acid fragment. Typically, the recombinant DNA construct portion of the construct includes a nucleic acid to be transcribed and translated, and a promoter. In the present invention, the recombinant DNA construct can be a plant artificial chromosome.

As used herein, "transgenic plant" includes reference to a plant modified by introduction of a heterologous nucleic acid.

As used herein, "telomere" refers to the end of a chromosome comprising a simple repeat DNA. The function of a telomere is to allow the ends of a linear DNA molecule to be replicated.

As used herein, "eukaryote" refers to living organisms whose cells contain nuclei. A eukaryote may be distinguished from a "prokaryote" which is an organism which lacks nuclei. Prokaryotes and eukaryotes differ fundamentally in the way their genetic information is organized, as well as their patterns of RNA and protein synthesis.

As used herein, "lower eukaryote" refers to a eukaryote characterized by a comparatively simple physiology and composition, and unicellularity. Examples of lower eukaryotes include flagellates, ciliates, and yeast.

As used herein, "higher eukaryote" refers to a multicellular eukaryote, characterized by its greater complex physiological mechanisms as well as its ability to interact with its environment in a more sophisticated manner. Generally, more complex organisms such as plants and animals are included in this category. Preferred higher eukaryotes to be transformed by the present invention include, for example, monocot and dicot angiosperm species, gymnosperm species, fern species, plant tissue culture cells of these species, and algal cells. It will of course be understood that prokaryotes and eukaryotes alike may be transformed by the methods of this invention.

As used herein, a "selectable marker" is a gene whose presence results in a clear pheontype, and most often a growth advantage for cells that contain the marker. This growth advantage may be present under standard conditions, altered conditions such as elevated temperature, on in presence of certain chemicals such as herbicides or antibiotics. Examples of selectable markers include the thymidine kinase gene, the cellular adenine-phosphoriboysltransferase gene and the dihydrylfolate reducast gene, hygromycin phosphotransferase genes, the bar gene and the neomycin phosphotransferase genes, among others. Preferred selectable markers in the present invention include genes whose expression confer antibiotic or herbicide resistance to the host cell, sufficient to enable the maintenance of a vector with a host cell, and which facilitate the manipulation of a plasmid into new host cells.

As used herein, "nucleotide" refers to one of the monomeric units from which DNA or RNA polymers are constructed, consisting of a purine or pyrimidine base, a pentose, and a phosphoric acid group. The nucleotides of DNA are deoxyadenylic acid, thymidylic acid, deoxyguanilic acid, and deoxycytidylic acid. The corresponding nucleotides of RNA are adenylic acid, uridylic acid, guanylic acid, and cytidylic acid.

# SEQUENCE LISTING

The present application also contains a sequence listing that contains 8 sequences. The sequence listing contains nucleotide sequences. For the nucleotide sequences, the base pairs are represented by the following base codes:

| Symbol | Meaning |
|---|---|
| A | A; adenine |
| C | C; cytosine |
| G | G; guanine |
| T | T; thymine |
| U | U; uracil |
| M | A or C |
| R | A or G |
| W | A or T/U |
| S | C or G |

| Symbol | Meaning |
|---|---|
| Y | C or T/U |
| K | G or T/U |
| V | A or C or G; not T/U |
| H | A or C or T/U; not G |
| D | A or G or T/U; not C |
| B | C or G or T/U; not A |
| N | (A or C or G or T/U) |

Nucleic Acid Sequences

In one embodiment, the present invention relates to isolated and purified nucleic acids which encode a functional centromere from *Oryza sativa*. As used herein, the term " a functional centromere" refers to the centromere or chromosome site that directs or supports kinetechore formation. The kinetochore is the physical structure that mediates the attachment of the spindle fibers to the chromosome and is therefore responsible for the proper partition of the chromosomes at mitosis and meiosis.

The nucleic acids of the present invention encode seven (7) different repetitive regions of centromeric DNA from *Oryza sativa*. Exemplary nucleic acids for such centromeres have the nucleotide sequences shown in SEQ ID NO:1, SEQ ID NO:2, SEQ ID NO:3, SEQ ID NO:4, SEQ ID NO:5, SEQ ID NO:6 and SEQ ID NO:7 or combinations thereof. SEQ ID NO:1 is also referred to herein as RCS1. SEQ ID NO:2 is also referred to herein as: RCS2. SEQ ID NO:3 is also referred to herein as RCH1. SEQ ID NO:4 is also referred to herein as RCH2. SEQ ID NO:5 is referred to herein as RCH3. SEQ ID NO:6 is referred to herein as RCE1. SEQ ID NO:7 is referred to herein as RCE2.

The present invention also contemplates nucleic acids which hybridize under stringent hybridization conditions to the nucleotide sequences set forth above. Generally, stringent conditions are selected to be about 5°C to about 20°C lower than the thermal melting point (Tm) for the specific sequence at a defined ionic strength and pH. The Tm is the temperature (under defined ionic strength and pH 7) at which 50% of the target sequence hybridizes to a perfectly matched probe. Typically, stringent wash conditions are those in which the salt concentration is about 0.22 molar at pH 7 and the temperature is at least about 50°C. However, nucleic acids which do not hybridize to each other under stringent conditions are still substantially identical if it encodes a substantially identical and functional centromere. This may occur, e.g., when a copy of a nucleic acid is created using the maximum codon degeneracy permitted by the genetic code. The present invention also contemplates naturally occurring allelic variations and mutations of the nucleotide sequences set forth above so long as those variations and mutations code, on expression, for a functional centromere.

As is well known in the art, because of the degeneracy of the genetic code, there are numerous other DNA and RNA molecules that can code for the same functional centromere as encoded by SEQ ID NOS: 1-7, or portions thereof. The present invention, therefore, contemplates those other DNA and RNA molecules, which, on expression, encode for a functional centromere encoded by the nucleic acid sequences of SEQ ID NOS: 1-7 or portions thereof. With knowledge of all triplet codons for each particular amino acid residue, it is

possible to describe all such encoding RNA and DNA sequences. DNA and RNA molecules other than those specifically disclosed herein and, which molecules are characterized simply by a change in a codon for a particular amino acid are within the scope of this invention. A table of codons representing particular amino acids is set forth below in Table 1.

**TABLE 1**

| First Position (5' end) | Second Position | | | | Third Position (3' end) |
|---|---|---|---|---|---|
| | T/U | C | A | G | |
| T/U | Phe | Ser | Tyr | Cys | T/U |
| | Phe | Ser | Tyr | Cys | C |
| | Leu | Ser | Stop | Stop | A |
| | Leu | Ser | Stop | Stop | G |
| C | Leu | Pro | His | Arg | T/U |
| | Leu | Pro | His | Arg | C |
| | Leu | Pro | Gln | Arg | A |
| | Leu | Pro | Gln | Arg | G |
| A | Ile | Thr | Asn | Ser | T/U |
| | Ile | Thr | Asn | Ser | C |
| | Ile | Thr | Lys | Arg | A |
| | Met | Thr | Lys | Arg | G |
| G | Val | Ala | Asp | Gly | T/U |
| | Val | Ala | Asp | Gly | C |
| | Val | Ala | Glu | Gly | A |
| | Val | Ala | Glu | Gly | G |

The nucleic acid sequences of the present invention can be used in marker-aided selection using techniques which are well-known in the art. Marker-aided selection does not require the complete sequence of the gene or precise knowledge of which sequence confers which specificity. Instead, partial sequences can be used as hybridization probes or as the basis for

oligonucleotide primers to amplify by PCR or other methods to identify nucleic acids specific for functional centromeric DNA in other plants and animals.

Plant Artificial Chromosome

In a second embodiment, the present invention relates to a plant artificial chromosome. More specifically, the nucleic acid sequences of the present invention can be used to construct a plant artificial chromosome vector. A plant artificial chromosome must contain the following essential elements: (1) autonomous replication sequences (hereinafter referred to as "ARS"), (2) a centromere which is functional in plants, and (3) telomeres which are functional in plants.

Autonomous Replication Sequences

ARSs have been isolated from the unicellular fungi *Saccharomyces cerevisiae* (brewer's yeast) and *Schizosaccharomyces pombe* (see Stinchcomb et al.. *Nature* 282:39- 43 (1979) and Hsiao et al., *J. Proc. Natl. Acad. Sci. USA* 76:3829-3833 (1979)). ARSs behave like replication origins allowing DNA molecules that contain the ARS to be replicated as an episome after introduction into the cell nuclei of these fungi. Although plasmids containing these sequences replicate, they do not segregate properly.

U.S. Patent 5,270,201 (hereinafter the " '201 Patent"), hereby incorporated by reference, discloses a method for isolating ARS sequences for use in higher eukaryotic organisms, by the formation of minichromosomes derivative of natural chromosomes. It has been demonstrated in yeast that inverted repeats of telomeric sequences are "resolved" by an unknown mechanism which results in a double-stranded cleavage between inverted repeats. After an inverted telomere repeat is introduced into a chromosome, a resolution reaction will lead to scission of the chromosome and formation of two chromosomal fragments, each with two telomeres. This process generates a minichromosome small enough to be isolated intact allowing further manipulation by *in vitro* techniques to delimit the sequences required for autonomous replication.

A second approach for an obtaining ARS is also disclosed in the '201 Patent. This approach is referred to as a "shotgun cloning approach". Higher eukaryotic organisms have many replication origins distributed throughout their genomes. For example, the *A. thaliana* genome contains approximately 1000 origins spaced every 70 kb along the chromosome. Therefore, the shotgun cloning approach involves looking for random fragments of genomic DNA throughout the genome of interest which promote extrachromosomal replication.

Autonomous replication sequences for use in the plant artificial chromosome of the present invention can be obtained using methods which are well known in the art. Autonomous replication sequences from yeast, such as those described above, can be used in the present invention. Moreover, ARS sequences from higher eukaryotic organisms obtained using the methods described in the '201 patent can also be used in plant artificial chromosome of the present invention.

Telomeres

Telomeres are believed to be involved in the priming of DNA replication at the chromosome end (see, Blackburn et al., *Ann. Rev. Biochem.* 53:163-194 (1984)). This is because conventional DNA polymerases are template dependent, synthesize DNA in the 5' to 3' direction, and require an oligonucleotide primer to donate a 3' OH group. When this primer is removed, unreplicated single-stranded gaps arise; most of these gaps can be filled in by priming from 3' OH groups donated by newly replicated strands located at the 5' end of the gap. However, the unreplicated gaps which lie next to the extreme 5' end of the DNA duplex cannot be primed in this manner. Consequently, telomeres must provide an alternative priming mechanism.

Telomeres are also responsible for the stability of chromosomal termini. Telomeres act as "caps," suppressing the recombinogenic properties of free, unmodified DNA ends (see Blackburn et al., *Ann. Rev. Biochem.* 53:163-194 (1984)). This reduces the formation of damaged and rearranged chromosomes which arise as a consequence of recombination-mediated chromosome

fusion events.

Telomeres may also contribute to the establishment or maintenance of intranuclear chromatin organization through their association with the nuclear envelope (see, for example, Fussell, C. P., *Genetica* 62:192-201 (1984)).

Telomeric or telomeric-like DNA sequences have been cloned from several lower eukaryotic organisms, principally protozoans and yeast. The ends of the *Tetrahymena* linear DNA plasmid have been shown to function like a telomere on linear plasmids in *S. cerevisiae* (see Szostak, J. W., Cold Spring Harbor Symp. *Quant. Biol.* 47:1187-1194 (1983)). A telomere from the flagellate *Trypanosoma* has been cloned (see, for example, Blackburn et al., *Cell* 36:447-457 (1984)). A yeast telomeric sequence has been identified (see, for example, Shampay et al., *Nature* 310:154-157 (1984)).

U.S. Patent 5,270,201 disclose a method for obtaining a telomere from a higher eukaryotic organism, specifically, from *Arabidopsis thaliana*. The telomeric sequences disclosed in the '201 Patent contain a tandem repeat of the sequence 5'-CCCTAAA-3.

Any telomeric sequence which produces a telomere which is functional in plants can be inserted into the plant artificial chromosome of the present invention. The telomeric sequence may be from yeast or from a higher eukaryotic organism as described above. Preferably, the plant artificial chromosome of the present invention will contain two (2) telomeric sequences.

Construction of a Plant Artificial Chromosome

Once the essential elements of a plant artificial chromosome are obtained (the ARS, centromere and teleomeres), a plant artificial chromosome vector can be constructed using methods which are well-known in the art (see, for example, Maniatis, T., et al., *Molecular Cloning: A Laboratory Manual* (Cold Spring Harbor, 1982)).

In addition to the essential elements described above, preferably positive and negative selectable plant markers (for example, antibiotic or herbicide resistance genes), and a cloning site for insertion of foreign DNA are preferably included. In order to propagate vectors in *E. coli*, it is necessary to convert the linear molecule into a circle by the addition of a stuffer fragment between the telomeres. In addition to the stuffer fragment, the artificial plant chromosome may also contain a origin of replication that can function in plants.

Artificial plant chromosomes which replicate in yeast also may be constructed to take advantage of the large insert capacity and stability of repetitive DNA inserts afforded by this system (Burke et al., *Science*, 236:806-812 (1987)). In this case, yeast ARS and centromere sequences are added to the artificial chromosome. The artificial chromosome is maintained in yeast as a circular molecule using a stuffer fragment to separate the teleomeres.

Nucleic acids for the essential components of the plant artificial chromosome obtained from any source whatsoever, may be purified and inserted into the plant artificial chromosome at any appropriate restriction endonuclease cleavage site. The nucleic acids usually will contain various regulatory signals (for example, promoters, termination segments, enhancers, etc., which are well known in the art) that allow for the expression of proteins encoded by the nucleic acids. Alternatively, regulatory signals residing in the artificial chromosome may be utilized.

The techniques and procedures required to accomplish insertion are well-known in the art (see Maniatis *et al.*, *Molecular Cloning: A Laboratory Manual*, Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y., (1982)). Typically, this is accomplished by incubating a circular plasmid or a linear DNA fragment in the presence of a restriction endonuclease such that the restriction endonuclease cleaves the DNA molecule. Endonucleases preferentially break the internal phosphodiester bonds of polynucleotide chains. They may be relatively unspecific, cutting polynucleotide bonds regardless of the surrounding nucleotide sequence. However, the endonucleases which cleave only a specific nucleotide sequence are called restriction enzymes. Restriction endonucleases generally internally cleave DNA molecules at specific recognition

sites, making breaks within "recognition" sequences that in many, but not all, cases exhibit two-fold symmetry around a given point. Such enzymes typically create double-stranded breaks.

Many of these enzymes make a staggered cleavage, yielding DNA fragments with protruding single-stranded 5' or 3' termini. Such ends are said to be "sticky" or "cohesive" because they will hydrogen bond to complementary 3' or 5' ends. As a result, the end of any DNA fragment produced by an enzyme, such as *Eco*RI, can anneal with any other fragment produced by that enzyme. This properly allows splicing of foreign genes into plasmids, for example. Some restriction endonucleases that may be particularly useful with the current invention include *Hind*III, *Pst*I, *Eco*RI, and *Bam*HI.

Some endonucleases create fragments that have blunt ends, that is, that lack any protruding single strands. An alternative way to create blunt ends is to use a restriction enzyme that leaves overhangs, but to fill in the overhangs with a polymerase, such as klenow, thereby resulting in blunt ends. When DNA has been cleaved with restriction enzymes that cut across both strands at the same position, blunt end ligation can be used to join the fragments directly together. The advantage of this technique is that any pair of ends may be joined together, irrespective of sequence.

Those nucleases that preferentially break off terminal nucleotides are referred to as exonucleases. For example, small deletions can be produced in any DNA molecule by treatment with an exonuclease which starts from each 3' end of the DNA and chews away single strands in a 3' to 5' direction, creating a population of DNA molecules with single-stranded fragments at each end, some containing terminal nucleotides. Similarly, exonucleases that digest DNA from the 5' end or enzymes that remove nucleotides from both strands have often been used. Some exonucleases which may be particularly useful in the present invention include *Bal*31, *S*1, and *Exo*III. These nucleolytic reactions can be controlled by varying the time of incubation, the temperature, and the enzyme concentration needed to make deletions. Phosphatases and kinases also may be used to control which fragments have ends which can be joined. Examples of useful

phosphatases include shrimp alkaline phosphatase and calf intestinal alkaline phosphatase. An example of a useful kinase is T4 polynucleotide kinase.

Once the source DNA sequences and vector sequences have been cleaved and modified to generate appropriate ends they are incubated together with enzymes capable of mediating the ligation of the two DNA molecules. Particularly useful enzymes for this purpose include T4 ligase, *E. coli* ligase, or other similar enzymes. The action of these enzymes results in the sealing of the linear DNA to produce a larger DNA molecule containing the desired fragment (see, for example, U.S. Pat. Nos. 4,237,224; 4,264,731; 4,273,875; 4,322,499 and 4,336,336, which are specifically incorporated herein by reference).

It is to be understood that the termini of the linearized plasmid and the termini of the DNA fragment being inserted must be complementary or blunt in order for the ligation reaction to be successful. Suitable complementarity can be achieved by choosing appropriate restriction endonucleases (*i.e.*, if the fragment is produced by the same restriction endonuclease or one that generates the same overhang as that used to linearize the plasmid, then the termini of both molecules will be complementary). As discussed previously, in a preferred embodiment, at least two classes of the vectors used in the present invention are adapted to receive the foreign oligonucleotide fragments in only one orientation. After joining the DNA segment to the vector, the resulting hybrid DNA can then be selected from among the large population of clones or libraries.

A method useful for the molecular cloning of DNA sequences includes *in vitro* joining of DNA segments, fragmented from a source high molecular weight genomic DNA, to vector DNA molecules capable of independent replication. The cloning vector may include plasmid DNA (see Cohen *et al.*, *Proc. Natl. Acad. Sci. USA*, 70:3240 (1973)), phage DNA (see Thomas *et al.*, *Proc. Natl. Acad. Sci. USA*, 71:4579 (1974)), SV40 DNA (see Nussbaum *et al.*, *Proc. Natl. Acad. Sci. USA*, 73:1068 (1976)), yeast DNA, *E. coli* DNA and most significantly, plant DNA.

A variety of processes are known which may be utilized to effect transformation; *i.e.*, the inserting of a heterologous DNA sequences into a host cell, whereby the host becomes capable of efficient expression of the inserted sequences.

Transformed Host Cells and Transgenic Plants

Methods and compositions for transforming a bacterium, a yeast cell, a plant cell, or an entire plant with one or more plant artificial chromosome vectors are further aspects of the present invention.

Means for transforming bacteria and yeast cells are well known in the art. Typically, means of transformation are similar to those well known means used to transform other bacteria or yeast such as *E. coli* or *Saccharomyces cerevisiae*. Methods for DNA transformation of plant cells include *Agrobacterium*-mediated plant transformation, protoplast transformation, gene transfer into pollen, injection into reproductive organs, injection into immature embryos and particle bombardment. There are various advantages and disadvantages associated with each of these methods.

Methods for transforming plant cells include any method by which DNA can be introduced into a cell, such as by *Agrobacterium* infection, direct delivery of DNA such as, by PEG-mediated transformation of protoplasts, by desiccation/inhibition-mediated DNA uptake, by electroporation, by agitation with silicon carbide fibers, by acceleration of DNA coated particles, etc.

Many methods for delivering genes into cells are known and well described. These methods include: (1) chemical methods (Graham et al., *Virology*, 54(2):536-539 (1973); Zatloukal et al. *Ann. N.Y. Acad. Sci.*, 660:136-153 (1992)); (2) physical methods such as microinjection (Capecchi, *Cell* 22(2):479-488 (1980)), electroporation (Wong et al., *Biochim. Biophys. Res. Commun.*, 107(2):584-587 (1982); Fromm et al., *Proc. Natl. Acad. Sci. USA*, 82(17):5824-5828 (1985); U.S. Patent 5,384,253) and microprojectile bombardment (i.e. the

gene gun) (Johnston et al., *Methods Cell. Biol.*, 43(A):353-365 (1994); Fynan et al., *Proc. Natl. Acad. Sci. USA*, 90(24):11478-11482 (1993); (3) viral vectors (Clapp, *Clin. Perionatol.*, 20(1):155-168 (1993); Lu et al., *J. Exp. Med.*, 178(6):2089-2096 (1993)); Eglitis et al. *Biotechniques*, 6(7):608-614 (1988); Eglitis et al. *Avd. Exp. Med. Biol.*, 241:19-27 (1988); and (4) receptor-mediated mechanisms (Curiel et al., *Proc. Natl. Acad. Sci. USA*, 88(19):8850-8854 (1991); Curiel et al., *Hum. Gen. Ther.*, 3(2):147-154 (1992); Wagner et al., *Proc. Natl. Acad. Sci. USA* 89(13):6099-6103 (1992)).

*Agrobacterium*-mediated transformation is a widely applicable system for introducing genes into plant cells because the DNA can be introduced into whole plant tissues. The use of *Agrobacterium*-mediated plant integrating vectors to introduce DNA into plant cells is well known in the art. Using conventional transformation vectors, chromosomal integration is required for stable inheritance of the foreign DNA. However, the artificial plant chromosome vector described herein may be used for transformation with or without integration, as the centromere function required for stable inheritance is encoded within the plant artificial chromosome. In particular embodiments, transformation events in which the plant artificial chromosome is not chromosomally integrated may be preferred, in that problems with site-specific variations in expression and insertional mutagenesis may be avoided.

The integration of the Ti-DNA is a relatively precise process resulting in few rearrangements. The region of DNA to be transferred is defined by the border sequences, and intervening DNA is usually inserted into the plant genome as described (Spielmann *et al.*, *Mol. Gen. Genet.*, 205:34 (1986); Jorgensen *et al.*, *Mol. Gen. Genet.*, (1987)). Modern *Agrobacterium* transformation vectors are capable of replication in *E. coli* as well as *Agrobacterium*, allowing for convenient manipulations as described (Klee *et al.*, *Bio/Technology*, 3:637-642(1985)). Moreover, recent technological advances in vectors for *Agrobacterium*-mediated gene transfer have improved the arrangement of genes and restriction sites in the vectors to facilitate construction of vectors capable of expressing various polypeptide coding genes. The vectors described (Rogers *et al.*, *Meth. In Enzymol.*, 153:253-277 (1987)), have convenient multi-linker

regions flanked by a promoter and a polyadenylation site for direct expression of inserted polypeptide coding genes and are suitable for present purposes. In addition, *Agrobacterium* containing both armed and disarmed Ti genes can be used for the transformations. In those plant strains where *Agrobacterium*-mediated transformation is efficient, it is the method of choice because of the facile and defined nature of the gene transfer.

*Agrobacterium*-mediated transformation of leaf disks and other tissues such as cotyledons and hypocotyls appears to be limited to plants that Agrobacterium naturally infects. *Agrobacterium*-mediated transformation is most efficient in dicotyledonous plants. Few monocots appear to be natural hosts for *Agrobacterium*, although transgenic plants have been produced in asparagus and more significantly in maize using *Agrobacterium* vectors as described (Bytebier *et al.*, *Proc. Natl. Acad. Sci. USA*, 84:5345 (1987)); U.S. Patent No. 5,591,616, specifically incorporated herein by reference). Therefore, commercially important cereal grains such as rice, corn, and wheat must usually be transformed using alternative methods. However, as mentioned above, the transformation of asparagus using *Agrobacterium* also can be achieved (see, for example, Bytebier *et al.*, *Proc. Natl. Acad. Sci. USA*, 84:5345 (1987)).

Other Transformation Methods

Transformation of plant protoplasts can be achieved using methods based on calcium phosphate precipitation, polyethylene glycol treatment, electroporation, and combinations of these treatments (see, for example, Potrykus *et al.*, *Mol. Gen. Genet.* 199:183 (1985); Lorz *et al.*, *Mol. Gen. Genet.*, 199:178 (1985); Fromm *et al.*, *Nature*, 312:791-793 (1986); Uchimiya *et al.*, *Mol. Gen. Genet.*, 204:204 (1986); Callis *et al.*, *Genes and Development*, 1:1183 (1987); Marcotte *et al.*, *Nature*, 335:454 (1988)).

Application of these systems to different plant strains for the purpose of making transgenic plants depends upon the ability to regenerate that particular plant strain from protoplasts. Illustrative methods for the regeneration of cereals from protoplasts are described

(Fujimura *et al.*, *Plant Tissue Culture Letters*, 2:74 (1985); Toriyama *et al.*, *Theor. Appl. Genet.*, 73:16 (1986); Yamada *et al.*, *Plant Cell Rep.*, 4:85 (1986); Abdullah *et al.*, *Biotechnology*, 4:1087 (1986)).

To transform plant strains that cannot be successfully regenerated from protoplasts, other ways to introduce DNA into intact cells or tissues can be utilized. For example, regeneration of cereals from immature embryos or explants can be effected as described (Vasil., *Biotechnology* 6:397 (1988)). In addition, "particle gun" or high-velocity microprojectile technology can be utilized (Vasil et al., *Biotechnology*, 10:667-674 (1992)).

Using that latter technology, DNA is carried through the cell wall and into the cytoplasm on the surface of small metal particles as described (Klein *et al.*, *Nature*, 327:70-73 (1987); Klein *et al.*, *Proc. Natl. Acad. Sci. USA*, 85:8502-8505 (1988); McCabe *et al.*, *Biotechnology*, 6:923 (1988)). The metal particles penetrate through several layers of cells and thus allow the transformation of cells within tissue explants.

By way of example, and not of limitation, Examples of the present invention will now be given.

## EXAMPLE 1: CLONING OF CENTROMERIC DNA FROM *ORZYA SATIVA*

### 1. Materials and Methods.

**a.** A rice BAC library was constructed from an indica rice (*Oryza sativa* ssp. Indica) line IR-BB21 and consisted of 11,000 clones (see Wang, G.-L., et al., *Plant J.* 7: 525-533 (1995), herein incorporated by reference). The cereal centromeric DNA element pSau3A9 (Jiang, J., et al., *Proc. Natl,. Acad. Sci. USA* 93:14210-14213 (1996), herein incorporated by reference) was used to isolate the rice centromere-specific BAC clones. The DNA sequence of pSau3A9 is shown in Figure 1 and SEQ ID NO:8. Rice lines used in this example include a javanica rice (*O. sativa* ssp. Javanica) line DV85, a japonica rice (*O. sativa* ssp. Japonica ) line Norin 28, and

indica rice line IR72, and four other *Oryza* species (*O. glaberrima, O. rufipogon, O. officinalis,* and *O. alta*). Gramineae species used in conservation studies include two species from the Bambusoideae subfamily [bamboo (*Bambusa vulgaris*), *Pharus* sp.], three species from the Panicoideae subfamily [sorghum, maize (*Zea mays*), and sugar cane (*Saccharum officinarum*)], six species form the Pooideae subfamily [*Agropyron intermedium*, barley (*Hordeum vulgare*), oats (*Avena sativa*), rye (*Secale cereale*), wheat (*Triticum aestivum*), and *Aegilops squarrosa*]. Three non-Gramineae species, *Juncus effusus*, *Cyperus alternifolius*, and *A. thaliana*, and rye and maize lines containing B chromosomes also were included.

**b. BAC Library Screening.** BAC filter preparations and BAC library screening were conducted as described in Wang, G.-L., et al., *Plant J.* 7:525-533 (1995); Hoheisel, J. D., et al., *Cell* 73:109-120 (1993), herein incorporated by reference. BAC clones were isolated by using pSau3A9 as a probe, and their cytological locations were confirmed by fluorescence *in situ* hybridization (hereinafter referred to as "FISH").

**c. Subcloning and Sequencing.** DNA fragments recovered from agarose gels were subcloned into pUC18 plasmids as described in Jiang, J., et al., *Proc. Natl,. Acad. Sci. USA* 93: 14210-14213 (1996), herein incorporated by reference. Cycle sequencing reactions were performed by using Applied Biosytems AmpliTaq DNA polymerase, FS Dye Terminator Ready Reactions kit, and a Perkin-Elmer Thermocycler (model 2400). Reaction products were analyzed on an Applied Biosystems DNA sequencer (model 373).

**d. Southern Blot Hybridization.** Plant genomic DNA was isolated as described in Gill, K. S., et al., *Genome* 34:362-374 (1991), herein incorporated by reference. BAC DNA was prepared by using an alkaline lysis method described in Sambrook, J., Fritsch, E. F. & Maniatis, T. (1989) *Molecular Cloning: A Laboratory Manual* (Cold Spring Harbor Lab. Press, Plainview, NY), 2nd Ed., pp. 1.25-1.26, herein incorporated by reference, and purified by CsCl ultracentrifugation. Gel transfers, prehybridizations, hybridizations, and posthybridization

washing were all as previously described Jiang, J., et al., *Proc. Natl. Acad. Sci. USA* 93: 14210-14213 (1996).

**c. Slot Blot Hybridization**. Copy number of each subclone in rice genome was determined by slot blot hybridization as described Zhao, X., et al., *Theor. Appl. Genet.* 78:201-209 (1989), herein incorporated by reference. Band intensities were measured on the autoradiographs by IPLab Spectrum v3.1 software.

**f. FISH**. Detailed protocols for FISH and Fiber-FISH are described in Jiang, J., et al., *Proc. Natl. Acad. Sci. USA* 93: 14210-14213 (1996) and Fransz, P. F., et al., *Plant J.* 9:421-430 (1996), herein incorporated by reference. The formamide in the hybridization mixture was 50% and 30% in regular and low stringency hybridizations, respectively. Washing was conducted at either low [2 × saline sodium citrate (SSC) at 42°C for 15 minutes], medium (50% formamide at 45°C for 15 minutes) or high stringency (70% formamide at 50°C for 15 minutes).

2. A rice BAC library constructed from indica rice (*Oryza sativa* ssp. Indica) and described by Wang, G.-L., et al., *Plant J.* 7: 525-533 (1995), was screened by using pSau3A9 as a probe. Twenty-two clones showed unambiguous positive hybridizations. Ten of the 22 clones were analyzed cytologically by FISH. Eight clones hybridized to the centromeric or/and paracentromeric regions of all rice chromosomes. Clone 17p22 showed bright and sharp signals specific to the centromeric regions. At a low hybridization stringency, this clone also hybridized exclusively to the centromeric regions of chromosomes from sorghum, maize, wheat, barley, oats and rye.

DNA from clone 17p22 was digested with restriction enzymes, *Bam*HI, *Dra*I, *Eco*RI, *Hae*III, *Hind*III, *Msp*I, *Pst*I, *Sau*3AI and *Sal*I and blotted onto nylon membrane. Small DNA fragments ranging from 0.5 to 3 kb were subcloned, and their distinctiveness was confirmed by Southern hybridization using blots containing 17p22 DNA digested with the above described nine restriction enzymes. Seven different DNA families, including two *Sau*3AI fragments

(subclones pRCS1 and pRCS2), three *Hind*III fragments (subclones pRCH1, pRCH2 and pRCH3), and two *Eco*RI fragment (subclones pRCE1 and pRCE2), were identified (see below in Table 2). These seven families hybridized to all of the fragments generated by the nine enzymes. FISH and Southern hybridization analysis indicated that all seven elements are repetitive in the rice genome (see below).

**Table 2**

**Summary of the Seven Rice Centromeric Repetitive DNA Families**

| Family | SEQ ID NO: | Size, bp | GC content, % | Organization pattern | Copy number* | Conservation |
|---|---|---|---|---|---|---|
| RCS1 | 1 | 1478 | 40 | Dispersed | 130 | Gramineae |
| RCS2 | 2 | 639 | 41 | Tandem | 6,200# | *Oryza* |
| RCH1 | 3 | 827 | 45 | Dispersed | 53 | Gramineae |
| RCH2 | 4 | 1,201 | 46 | Dispersed | 99 | Gramineae |
| RCH3 | 5 | 1,341 | 48 | Dispersed | 67 | Gramineae |
| RCE1 | 6 | 701 | 39 | Dispersed | 287 | Bambusoideae |
| RCE2 | 7 | 2,863 | 41 | Dispersed | 305 | Gramineae |

*Based on the haploid genome of rice as 424 Mb (24).

#The copy number of the 168-bp monomer in the rice genome.

Clone pRCS1 contains a 877-bp *Sau*3AI fragment that hybridizes to the pSau3A9 sequence. Sequencing analysis revealed that the 259 bp at the 3' end of pRCS1 had 80% sequence identity to the central part (bases 338-602) of the pSau3A9 sequence (see Jiang, J., et al., *Proc. Natl,. Acad. Sci. USA* 93: 14210-14213 (1996), FIG. 1 and SEQ ID NO:8). The first 95 bp in pRCS1 had 76% sequence identity to a Ty3/*gypsy* class of retrotransposon sequence reported in maize (GenBank Accession No. AF030633). Nucleotides 171-228 of pRCS1 had 70% sequence identity to a Ty3/*gypsy* class of retrotransposon sequence reported in *Lilium*

*henryi* (X13886). It was also discovered that the pSau3A9 sequence in sorghum has similar sequence identities to the Ty3/*gypsy* class of retrotransposons. These results indicated that both pSau3A9 and pRCS1 probably were derived from retrotransposon-related DNA sequences.

The RCS1 sequence was located in the centromeric regions of all 24 rice chromosomes by FISH (see FIG. 2*A*). The sizes and intensities of the FISH signals were uniform on different chromosomes, suggesting that all rice chromosomes contain a similar number of copies of this element. Slot blot analysis suggested that ther are about 130 copies of RCS1 present in the haploid genome of japonica rice DV85 (see Table 2).

Rice genomic DNA was digested with several restriction enzymes and probed with the 259-bp fragment conserved between rice and sorghum. One or few major bands and several minor bands were detected in most of the lanes (see FIG. 3). Fiber-FISH using pRCS1 as a probe did not generate clustered signals. These results suggested that the RCS1 sequence is dispersed in the centromeric regions of rice chromosomes.

FISH analysis revealed that pRCS1 also hybridized exclusively to centromeric regions of chromosomes from other *Gramineae* species (see FIG. 2*B-E*). The FISH results on rye (see FIG. 2*B*) and barley (see FIG. 2*C*) chromosomes showed that hybridization was exclusive to the primary constrictions. FISH signals also were detected in the centromeres of the supernumerary B chromosomes from both rye and maize (see FIG. 2 *B* and *E*). Positive Southern hybridization signals were detected in all other Gramineae species analyzed, including bamboo, *Pharus* sp., oats, wheat, sugar cane, *Ae. squarrosa*, and *Ag. intermedium*. However, homologous sequences could not be detected by Southern hybridization analysis in dicot species and any monocot species outside of Gramineae, indicating that the RCS1 family is sufficiently conserved only in the grass family Gramineae.

Clone pRCS2 contains a 639-bp *Sau*3AI fragment consisting of four copies of a tandemly arranged repeat with a consensus sequence of 168bp (see FIG. 4). The four copies were 84-91%

identical with one another. The third copy of the repeat contains a 6-bp insertion (TTGGCC) at base 147. A search of the GenBank database found a highly significant match to a repetitive DNA element isolated from *O. Sativa* (GenBank Accession No. U63977).

Southern hybridization analysis of rice genomic DNA using probe pRCS2 revealed ladder patterns using several restriction enzymes, including *Dpn*II, *Sau*3AI, *Msp*I, *Hpa*II, and *Hae*III. indicating that the RCS2 family is tandemly arranged in the rice genome (see FIG. 5). Several enzymes produced digestion profiles comprised of monomer and multiples (dimer, trimer, tetramer, etc.) of the 168-bp basic repeat.

Probe pRCS2 hybridized only to the centromeric regions on all rice chromosomes (see FIG. 2*K*). Significant variation in the size and intensity of the FISH signals was detected in different centromeres. Two pairs of chromosomes had strong signals, and a third pair had very faint signals (see FIG. 2*K*). All of the signals became weaker as the posthybridization washing stringency was increased (see FIG. 2*L*). However, even after washing in 70% formamide at 50°C for 15 minutes, most signals were still discernible (see FIG. 2*M*), suggesting that the signal disparity reflects difference in copy numbers rather than sequence divergence of the RCS2 family in different rice centromeres. Though the longest chromosome (chromosome 1) had the strongest signals, it was not possible to relate the copy numbers to the chromosome sizes. It was evident that the weakest signals were not on the smallest chromosomes (see FIG. 2*K*).

Three subspecies of *O. sativa* (AA genome), together with *O. glaberrima* (AA), *O. rufipogon* (AA), *O. alta* (CCDD), and *O. officinalis* (CC) were included for FISH analysis. FISH signals were detected in the centromeric regions from all of the chromosomes of these species. Southern hybridization analysis revealed that the RCS2 family is present only in the species within genus *Oryza*. Homologous sequences could not be detected even at a low stringency in any plant species outside of genus *Orya*.

RCS2 is the most abundant element isolated from BAC 17p22 and has about 1,550 copies, corresponding to 6,200 monomers, in the haploid genome of DV85 (see Table 2 above). BAC 17p22 contains about 46 copies of this element, corresponding to approximately 39% of the BAC insert. Fiber-FISH analysis demonstrated that the RCS2 family is organized into various sizes of uninterrupted arrays in the rice genome. The longest observed block with small interspersed gaps (<2 $\mu$m) was 51 $\mu$m (see FIG. 2*N*). Based on a 2.96-kb/$\mu$m resolution of the Fiber-FISH technique, this block represents approximately 151 kb of uninterrupted RCS2 sequences. The longest observed single Fiber-FISH signal with interspersed gaps larger than 2 $\mu$m was 188 $\mu$m representing approximately 556 kb of centromeric DNA sequences.

The other five centromeric DNA elements isolated from rice BAC clone 17p22 were analyzed by FISH, and all of them hybridized exclusively to the centromeric regions of all rice chromosomes (see FIG. 2*F-J*). One or two pairs of rice metaphase chromosomes showed weak hybridization when pRCH2, pRCH3, and pRCE2 were used as probed. No relationship can be confirmed between signal intensities and the sizes of the chromosomes.

The sequence information for these families is listed in Table 2, above. Searching in the GenBank database did not uncover any significant matches to these sequences except for pRCH2. Bases 39-102 and 204-232 in pRCH2 had sequence identities to the centromeric CCS1 sequence isolated from *B. sylvaticum* (see Aragon-Alcaide, L., et al., *Chromosoma* 105, 261-268 (1996); Abbo, S., et al., *Chromosome Res.* 3:5-15 (1995)). Interestingly, about 120 bp (bases 8-130) of this element had 80% sequence identity to the spacer sequence that separates the rice 5S rRNA genes. The possibility that this element associates with the 5S rDNA locus was excluded because the FISH signals from pRCH2 was located proximal to those from the 5S rDNA locus.

In Southern hybridization analysis, all five elements produced one or few major bands and several minor bands under several restriction enzymes, similar to the RCS1 family (see FIG. 3), suggesting that they all are dispersed in the rice centromeric regions. The copy numbers of these elements ranged from 53 to 305 copies per haploid rice genome (see Table 2, above).

All five elements were hybridized to various plant species by Southern hybridization. The RCE1 family was present only in the species from the Bambusoideae subfamily, including rice, bamboo, and *Pharus* sp. (See FIG. 6*B*), whereas RCH1, RCH2, RCH3, and RCE2 all were conserved across the Gramineae species (see FIG. 6*A* for RCH1). Species from subfamily Panicoideae and Bambusoideae generally had stronger hybridization signals than those from subfamily Pooideae (see FIG. 6*A*).

The cytosine nucleotides, especially those in dinucleotide sequence 5'CpG3', are the most common sites for methylation in plant genomes. Methylation occurs at lower frequencies when the C and G are separated by 1-2 A/T nucleotides (see Gruenbaum, T., et al. *Nature* (*London*) 292: 860-862 (1981)). Enzymes *Msp*I and *Hpa*II ar isoschizomers that recognize the 5'CCGG3' sequence. Neither enzyme can cut when the 5'C is methylated, and only *Msp*I can cleave when the internal cytosine is methylated. Though both enzymes produced similar digestion profiles of rice genomic DNA, *Msp*I generated much smaller-sized hybridization bands from all of the rice centromeric DNA probes than *Hpa*II did (see FIG. 3 for RCS1 and FIG. 5 for RCS2). For the RCS2 element, monomers of the 168-bp basic repeat could be found in *Msp*I lane, and most of the hybridization was in the fragments smaller than 2 kb, whereas the majority of hybridization in the *Hpa*II lane was larger than 2 kb (see FIG. 5). For the other centromeric elements, DNA fragments smaller than 5 kb were not detected in *Hpa*II lanes (see FIG. 3 for RCS1). These results suggest that the cytosine of the CpG dinucleotides are heavily methylated in the rice centromeric DNA sequences. Restriction enzyme *Sal*I recognizes 5'GTCGAC3' and is sentsitive to the methylation of CpG dinucleotides. Small fragments (<10 kb) that hybridized to the centromeric elements were not detected in the *Sal*I lanes (see FIGS. 3 and 5).

What is Claimed Is:

1. An isolated and purified nucleic acid comprising a nucleotide sequence of SEQ ID NO:7.

2. A recombinant DNA construct comprising a centromere, wherein said centromere comprises a number of highly repetitive regions of DNA having a nucleotide sequence of SEQ ID NO:7.

3. The recombinant DNA construct of claim 2, further comprising a yeast autonomous replication sequence.

4. The recombinant DNA construct of claim 2, further comprising an autonomous replication sequence from a higher eukaryotic organism.

5. The recombinant DNA construct of claim 2, further comprising a yeast telomere.

6. The recombinant DNA construct of claim 2, further comprising a telomere from a higher eukaryotic organism.

7. The recombinant DNA construct of claim 2, further comprising a selectable marker gene.

8. A plasmid comprising the DNA construct of claim 2.

9. The plasmid of claim 8, wherein said plasmid further comprises an origin of replication and a selectable marker that functions in bacteria.

10. The plasmid of claim 9, wherein said bacteria is *E. coli*.

11. The plasmid of claim 8, wherein said plasmid further comprises an origin of replication and a selectable marker that functions in yeast.

12. The plasmid of claim 8, wherein said yeast is *S. cerevisiae.*

13. A plant artificial chromosome vector comprising an autonomous replication sequence, two telomere sequences, a centromere sequence having the nucleotide sequence of SEQ ID NO:1, SEQ ID NO:2, SEQ ID NO:3, SEQ ID NO:4, SEQ ID NO:5, SEQ ID NO:6 and SEQ ID NO:7 or combinations thereof, and at least one selectable marker sequence.

14. The plant artificial chromosome vector of claim 13, wherein the autonomous replication sequence if from yeast.

15. The plant artificial chromosome vector of claim 13, wherein the autonomous replication sequence is from a higher eukaryotic organism.

16. The plant artificial chromosome vector of claim 13, wherein the telomere sequences are from a higher eukaryotic organism.

17. The plant artificial chromosome of claim 16, wherein the telomere sequences are from *Arabieopsis thaliana.*

18. The plant artificial chromosome vector of claim 13, wherein the telomere sequences are from yeast.

19. A plant cell transformed with the plant artificial chromosome vector of claim 13.

20. The transformed plant cell of claim 19, wherein the plant cell is from *Oryza sativa.*

21. A transgenic plant comprising the transformed plant cell of claim 19.

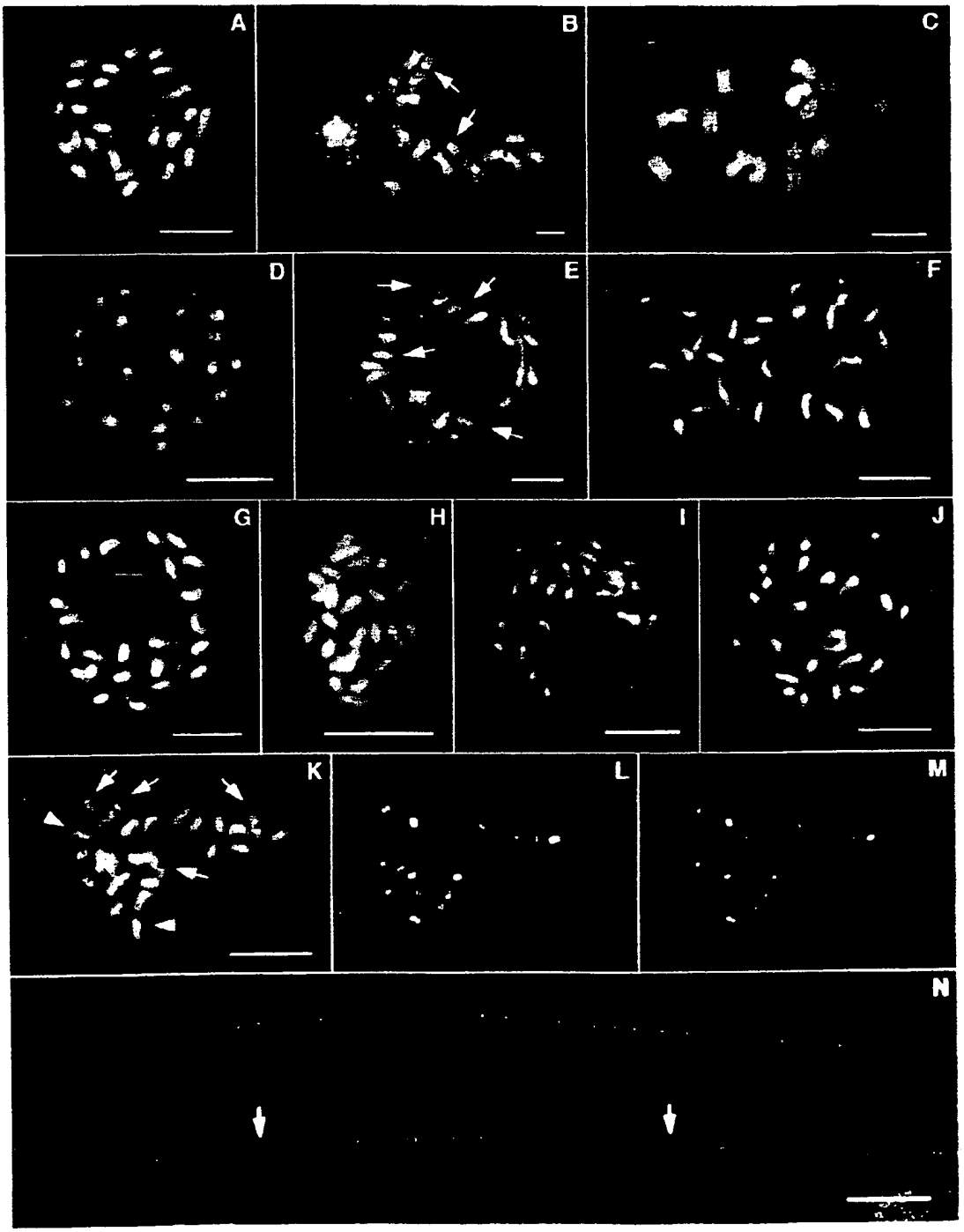
22. A method of identifying centromeric DNA in a higher eukaryotic organism, the method comprising the steps of:

hybridizing an isolated nucleic acid selected from the group consisting of SEQ ID NO:1, SEQ ID NO:2, SEQ ID:3, SEQ ID NO:4, SEQ ID NO:5, SEQ ID NO:6, SEQ ID NO:7 and combinations thereof, with a sample of DNA from a higher eukaryotic organism; and identifying and isolating centromeric DNA from said sample.

FIG. 1

```
        10         20         30         40         50
GATCTTTGGA TTGGAAACAG TTAAAGAACA ATATGTGCAT GATGATGATT

        60         70         80         90        100
TTAAAGATGT GTTTTTGCAT TGTAAGGATG GGAAGGCATG GAATAAATTT

       110        120        130        140        150
GTTGTAAATG ATGGTTTTGT GTTTAGAGCT AATAAGCTAT GCATTCCAGC

       160        170        180        190        200
TAGCTCTGTT CGTTTGTTGT TGCTACAGGA AGCACATGGA GGTGGTTTGA

       210        220        230        240        250
TGGGACATTT TGGGGCAAAG AAGACGGAGG ACATACTGGC TGGTCATTTC

       260        270        280        290        300
TTTTGGCCAA AGATGAGGAG AGATGTGGAG AGATTTATTG CTCGCTGCAC

       310        320        330        340        350
GACATGTCAA AAGGCCAAGT CACGCTTAAA TCCACACGAT TTGAAGCCAT

       360        370        380        390        400
ATTTGGGTGA GGGAGATGAG CTTGAGTCGG GGACGACTCA AATGCAAGAA

       410        420        430        440        450
GGGGAGGATG ATGAGGACAT CAGCACCATC TATACATCCA CACCTACACC

       460        470        480        490        500
CACACCATCG CCAACACCAC TTGGCCCTCT TACTCGTGCC AGTGCCCGTC

       510        520        530        540        550
AACTGAACCA TCAAGTAAGT TTATTCTTAA ACTCTTGTCC ATCATATTTA

       560        570        580        590        600
GACAATGGAG ACACGTGCAC TCTTGTTTTG CTTAGGAATG ATGGAGAGGA

       610        620        630        640        650
CCAGAAGCAT AGGGGATTGG TGTAGGCTGG ATTTGGACAG CAAGACAGCA

       660        670        680        690        700
CCAACTTACA ACAACCGCCA TGACTTCATA CAGAGTCCAT TTTAAGCATG

       710        720        730        740        750
CAAGCACTTG ATGGAAAACT CGTCAAGTAT ATTTTTAGAT GGATC
```

Fig. 2



# Fig. 3

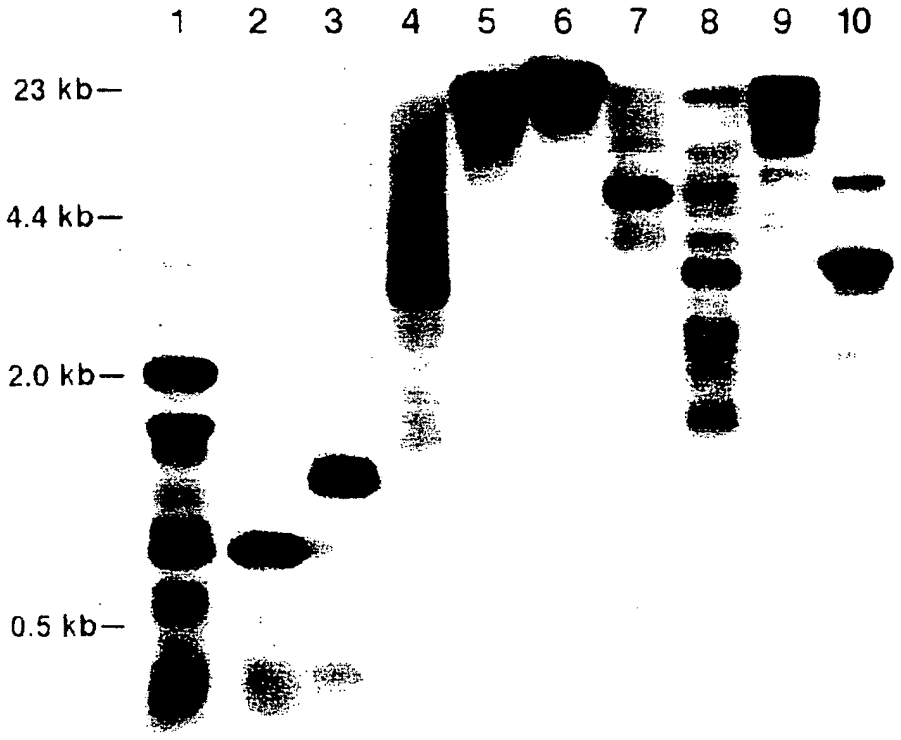
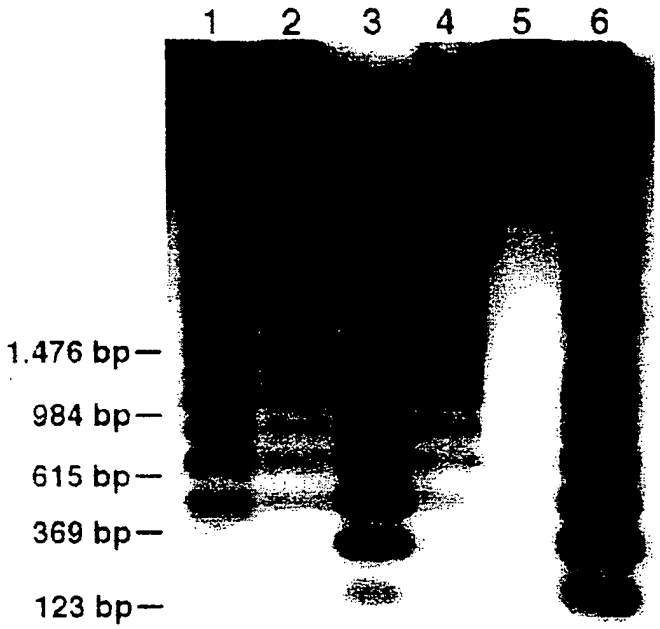


FIG. 4

```
      1                                                            50
A GATCTT-TTCTACT -GGAATCAAA ATGTTCAAAA AATGCCAAAA CATGATTTTT
B     TT-TTCTACT -GGAATCAAA ATGTTCAAAA AGAGCCAAAA CATGATTTTT
C     TTATTCTACT -GGAATCAAA ATGTTCAAAA AGAGCCAAAA CATGATTTTT
D     TTA-T-TAAT CGGAATCAAA ATGTTCAAAA GGCACCAAAA CATGATTTTT
F     TTatTcTAcT -GGAATcAAA AaGTTCAAAA agagCCAAAA CATGATTTTT

      51                                                           100
A     GGACATATTG GAGTGTATTG GGTGCATTCA TGGCAAAAAC TCACTCCGTG
B     GGACATATTG GAGTGTATTG GGTGCGTTCA TGGCAAAA-C TCACTTCGTG
C     GGACATATTG GAGTGTATTG GGTGCGTTCG TGGCAAAA-C TCACTTCGTG
D     TGACATATTG GAGTGTATTG GGTGCGTTCG TGGCAAAAAC TCACTTCGTG
F     gGACATATTG GAGTGTATTG GGTGCgTTCg TGGcAAAAaC TCACTtCGTG

      101                                                          150
A     ATTCGCGCGG CGAACTTTTG TCAATTAATG CCAATAT-TG GG-ACA----
B     ATTCGCGCGG CGAACTTTTG TCATTTAATG CCAATATGTG CATACA----
C     ATTCGCGCGG CGAACTTTTG TCAATTAATT CCAATATGTG CATATTTTGG
D     ATTCGCGCGG CGAACTTTTG TCATTTAATG CCAATAT-TG GC-ACA---G
F     ATTCGCGCGG CGAACTTTTG TCA*TTAATg CCAATATgTG *atAca---g

      151              168
A     --CGAG-G-G T-GCGATG  (155 bp)
B     --CGAGAGAG T-GCGATG  (158 bp)
C     CCCAAA-GTG TTGCGATG  (165 bp(
D     --CGA-CGGG T-GCGATC  (157 bp)
F     --CgAg*G*G T GCGATg
```

# Fig. 5



# Fig. 6

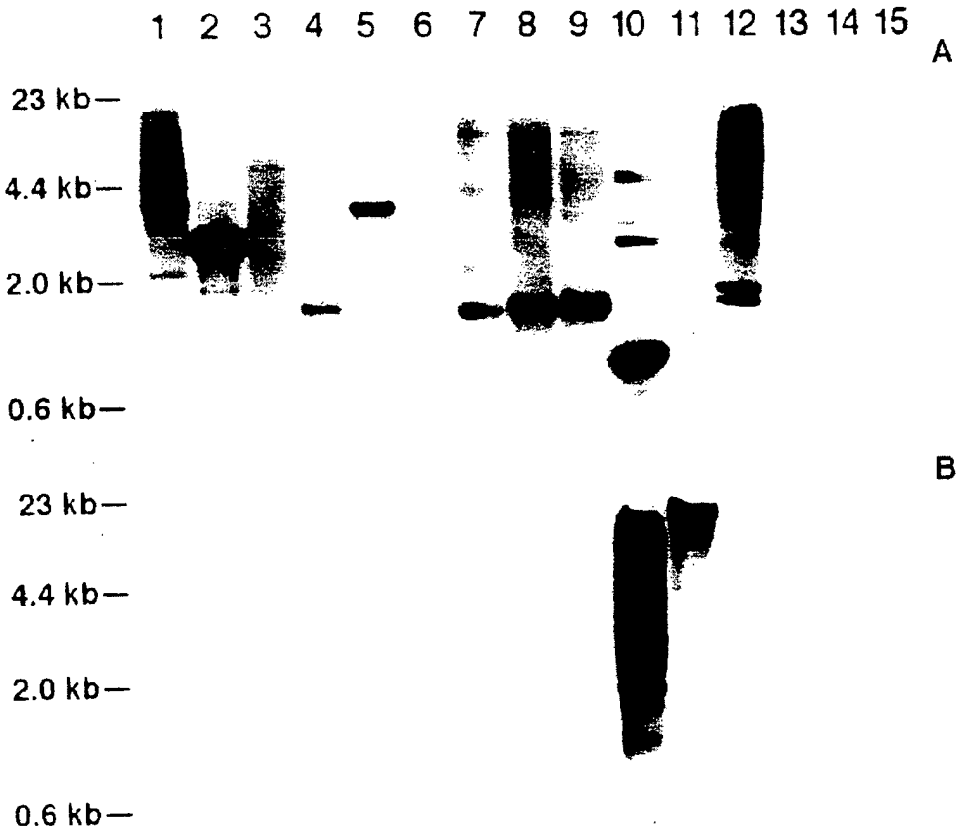

1 2 3 4 5 6 7 8 9 10 11 12 13 14 15
A
23 kb—
4.4 kb—
2.0 kb—
0.6 kb—
B
23 kb—
4.4 kb—
2.0 kb—
0.6 kb—

FIG. 7

```
1     tcatcaatga tgggtttgtt ttcagagcta acaagctatg cattccagct agctccgttc
61    gcttgtttgt gttgcaggaa gcgcatggag gcgatttgat gggacatttt ggtgccaaga
121   agacacatga catccttgct agtcatttct tctggccaca gatgcgaaga gatgttggca
181   ggttcgttgc tcgctacgct acatgtcaaa aggctaagtc acgcttacat ccacatggtt
241   tgtatatgcc tcttcctgtt cctactgttc cttgggaaga tatttcaatg gatttgtgt
301   taggattgcc taggaccaag aggggcgtg atagcatttt tgtggttgtg gatcgatttt
361   ctaaaatggc acatttcata ccatgtcata aaactgatga tgcttctcat atcgcttgat
421   tgttcttttcg agaaattgtt cgcttgcatg gtgtgccaaa cacaattgtt tctgatcgtg
481   acacaaaatt tcttagccat tttggagaga ctttgtgggc taaattgggg actaaacttt
541   tgttttctac tacttgtcat ccccaaactg atggacaaac tgaagtgggtg aatagaacct
601   tgtctactat gctttgaagaa aaatatcaag atgtgggaag aatgcttgcc
661   tcatattgaa ttgcttcctt gcattctact acaaaaaatg tgcccatttc
721   agattgtgta tggtttgttg cctcgtgctc caattgattt gatgccttta ccatcttctg
781   agaaactgaa tttgatgcg aagcaacgtg ctgagttgat gttaaaactg catgagacaa
841   ctaaagaaaa catagagcgc atgaatgcta agtataagtt tgctggtgac aaaggtagaa
901   gggaatttgaa ttttgaacct ggagatttgg ttttgttgca tttgcgaaaa gaacgatttc
961   ctgatttgag aaagtctaaa tagatgccta gagctgatgg accatttaaa gtgttagcaa
1021  agattaatga gaatgcatat aagattgatt tgcttgcaga tttcgtgggtt agtcccacat
1081  ttaacatttgc agatttgaag ccgtatattg ggagaaaaag atgagcttga gtcgaggatg
1141  actcaaatgc aagatgtggga ggatgatgag gacatcaaca ccatcgatac atccacgtcc
1201  ccccatatac agcatgatgg tcctattacc cgcgcttgtg cacgtcaact aaattatcag
1261  gtgattcttt cttgagttca aatttcctcg tctttatacc tcggagacgc gtgcactcgt
1321  gttttactcc aggaacgtat ggagaggatc aaagggaaga ggattcgcgc ggggtggatt
1381  cggactgcag ggcagcgcca acttctgacg gccgccacga cttcatgcaa actccgattt
1441  gggcgtgcaa gtacttcatg gaaagcttat caagtcta
```

FIG. 8

```
1    gatcttttct actggaatca aaatgttcaa aaaatgccaa aacatgattt ttggacatat
61   tggagtgtat tgggtgcatt catggcaaaa actcactccg tgattcgcgc ggcgaacttt
121  tgtcaattaa tgccaatatt gggacacgag ggtgcgatgt tttctactgg aatcaaaaag
181  ttcaaaaaga gccaaaacat gatttttgga catattgggg tgtattgggg tgcgttcgtg
241  gcaaaactca cttcgtgatt cgcgcggcga acttttgtca tttaatgcca atatgtgcat
301  acacgagaga gtgcgatgtt attctactgg aatcaaaaag ttcaaaaaga gccaaaacat
361  gatttttgga catattgggt tgtattgggt gcgttcgtgg caaaactcac ttcgtgattc
421  gcgcggcgaa cttttgtcaa ttaattccaa tatgtgcata ttttggccca aagtgttgcg
481  atgttattaa tcggaatgaa aaagttcaaa aggcaccaaa acatgatttt ttgacatatt
541  ggagtgtatt gggtgcgttc gtgggaaaaa ctcacttcgt gattcgcgcg gcgaacttct
601  gtcatttaat gccaatattg gcacagcgac gggtgcgat
```

FIG. 9

```
1    aagcttcata ctttgtccgt ttgaaataag attttctgtg gttggttcta gtgtgctcct
61   ggttgtgaaa aaaaatatca aaaaaataat aaaaaaaata aaaaaaagtt gtgcctctct
121  cctcagcact agagagagca agcaaaaaaa aaaaaagagg atccagcgag cacaccacac
181  catccccccc ttgtcaccac cttccaccac cggttctttgt tttcgttgct gtttgggttt
241  gtgttttgtg gcgtcctttg tggttagact cgtcgtctct acttcgacta gcctaggacc
301  agctactgta cactataccc tctgagcgat tattcaattt gctttggcta acgtggtttc
361  taatcttttt cttgagcaat gctttcagcc caccaacagc tccaccacaa acccgactac
421  agcttgacag gtcttgttgc tacagcaacg atacacctcg ttccacggta gctaactagt
481  ggtgttgctg tcccctacct gtgggcaagg taagaactgg taagagcttg tgagacaagc
541  tgcgagtgaa gtgagagtga gcgtcttgca aaagctacat cccaatagtt gtgtagggct
601  tacattacat cctccattgt tttgttgtcc ttcatttcta aaccatggca ggtccagagg
661  ataaagatga gtagggtgca tctcattccc cacgcaccaa aggcatcatc caatacttta
721  caaggcaagt gaagcagcac acagaaggac ttgatacaga tttgcaggtg acaaatgaga
781  agattggaca attggagtcc acgcagatct ccaccaacac taagctt
```

FIG. 10

```
1     aagcttctcg ccacaaaaac agaattttgc aagaggaatt tggcgatttt aaatttgccg
61    aacccagcaa agctggggag aaacggatgt aacttttttt gtagatgtcc gtggatatat
121   tataagccta atttcgagtc cgtatgaaaa agatatgact gttttaagga aggtcactcg
181   aatccagggtt ctatttcctt tctagtggcg gcaggataca aggcaactca caaaagctat
241   cggcaacaga tcaaacacaa atagaaagtg atttggtggg tgactgcgac ggtggtgacg
301   atggtgatga cggcaaaatt gatgaagtgg cgatgtgcgg tggtggcaag aactcgaaac
361   tctaatcgcc taacactaag accagcaaat taataaaacc gatgcaatcg aaaactcaac
421   aagccttaac taagcagtac taatgaagag gcacaatggt atatatgatc accgaaaaac
481   taatctacta ttttttgtt gttgtctttt ctggactata ggaaaaaaaa ataatgacga
541   agggaagggt aaattctctc accgatgaac cacgctctgc taccacctga tgcgaacccg
601   ctagggtttg tggcccgatc ttcgatgag aggtaaggga taactcgatt agatgaagtc
661   gacgttcacg gcccgactac aactgtccaa agacgctgtg ccttagcaac cgatacaccg
721   tctcccaatgg tcgccgcgaa cttgtggtgc gcgtcaaccc ggccacgagg gcactcgtcc
781   tgcaagcaat cgaagaacaa gcaagaacaa gtaggacaag cactgaaatt gctagataaa
841   gatgaaagtt tcactatcaa acacaatatg gtgggggttcc gattacaggc agacgcggcg
901   gtttagccaa cacgcgcgct gcgagccggt agcaagaagc tatcttctat catcaaaacc
961   cgcctgtttt tggcggcgac tagaggtata acacaagagg ggaagaacga ccataggctc
1021  gtgctccaac cctaggacgc gcccctaatg ggcccaacat ggatacacag cccattgggc
1081  caaaagaggt gacgcagcac catggacaga aaatagtcgg gagtaaaatg acaattgcgg
1141  ccgctccaga acagatatgg acatgtggct ggatccactt gaaagtagac ttgataagct
1201  t
```

FIG. 11

```
1     aagcttgtcg cccgcggcca ctttagnctt gctgctcttg tgacgcattt tgatgctctc
61    aatgctggag gcaatggtgg gggtaatgat gatgacattg acggagagta cgtcgaggat
121   aatttggagg atgaatatat tgctgacact gaacaagatg atcgagatgc tcgagatcgt
181   cgacggctac acaacaatcg acgaggtatg ggtggtcgcc gccgacgcga ggtacgcaac
241   aatgatgatg cttttctaa aattaaattt aagattcctc ctttgatgg aaaatatgat
301   cctgatgcgt acctctcttg ggagattgct gttgaccaaa aatttgcatg ccatgagttt
361   cctgagagca ctagagttag agctgctacg agtgagttca ccgatttttgc tttggtttgg
421   tggatagaac atggcaagaa aaatcctaat aacatgccac aaacttggga tgctttgaaa
481   cgggtcatgc gggctagatt tgttcctttct tactatgcac gcgacttgct gaataggttg
541   caacaattga gacagggtcg gaaaagtgta gaggaatatt atcaggagtt acaaatgggc
601   ttgcttcgtt gtaatttaga ggaaactgag gacgctgcca tggctagatt tttgggtggg
661   ttaaaccgcg agatttatga catcgtagac tataaagatt acgctaatat gacccgattg
721   tttcatttg ctgtaaggc tgaaagggaa gtgcaaggac gacgtgctag tgccaaggct
781   aatttttctg caggtaaaac ttcatcatgg cagacacgca ccactcctcc ggccggccgt
841   actgcttctc catcttccac acccacaacc agtcgagcag cacctcctcc atctagtgac
901   aagtcagcga caaaggctgc tcagccagca ccgagtgctt cttcaatggc atccacaggc
961   cgaatgagag atgttcagtg ccaccgttgc aagggctttg ggcatgtgca gcgtgactgc
1021  cctagcaagc gagttttggt agtcaaaaac gatggtgagt actcctctgc tagtgatttc
1081  gatgatgata cacttgcttt gcttgcggct gaccatgcag ataatgagcc accggaagag
1141  cacattgggg ctgcatttgc ggatcactat gagagcctca ttgtgcagcg tgtccttagc
1201  gcacaaatgg agaaggcgga gcaaaatcag cgacacacgt tgttccaaac aaagtgtgtc
1261  gtcaaagagc gttgttgccg catgatcatt gatggaggta gctgcaacaa cttggctagc
1321  agcgagatgg tggagaagct t
```

FIG. 12

```
1   gaattcctct gtcacaagtt gatttactcg ctgttccttg tgataaagaa gagttgtgtg
61  ataatgcttc acttaaatcc atgccacaac tagtgaatga acatgctatt cctagtgtat
121 ctttctgtgc tgatttaag catgttgttc acattgctaa tagagtagag gaacgtgaat
181 tgacatcttc tttaaatact ttgggctatg ttcagtttgc tgattttcat gagctcgata
241 atttgaagga gaaattattt gctaagtctg atttgccatg tccaagtaac gctatttttc
301 atctctttgg tgaatataat gatagaggaa tatatttggt gcatagagtt tacatctgtt
361 cagatttaga acctcctgta catgtggata aaacatgcaa gctagagaga aatgttattg
421 ctaacaaaat tgtcttcgagt ttgtcttgtt ttgattggac gaaacaggtt gttgtagaat
481 ggtactacgc aatagagcac cactatggag ataataccga ggacggtttt ccattgaaga
541 tacgggagg tatgatgtga ccatggctac acacggatgc aactcatgtg actcacatgc
601 ataggatgaa cagagaagat atcgagatca catcgtccaa gtgctggaat ccgattcggc
661 cacactacta ctactgctga cttcaaacgg ccgccgaatt c
```

# FIG. 13

GAATTCCATCTCGTGTGGCCGGAATCACAGGCACAAGGTGCTGCGTCACCTATA
ATGGGCCTATGGGCTTGTAACTTCATTTGGGACCCCGGCCCAGGTGGGGCCCA
TGTGGGGTGCGCCCCAACCAGGTGGAGCACGACCCTAGGCACCCCTAGGTCGT
CCTCCCACCCCTATATATAGCTAGGTACCCCTTCAGGGTTTCTTGGGTTTTGATA
GATTAAAGTTTAGCCATTGCTACTTGCTTGCAGCGCGCGTGTCGCCTAGACCGT
CCGTCTGCTTGTTCTTCGGAACCCCAACTTATCATTTGTATTAAATTCCTATTTG
CAATATCAGATTGCTTTTATCTTGTTCTTGCTTGTTTCTTCGATTTGCTTGCAGG
AATAGGGTTGATCTGCACCGGTAAGATCAACAACCCACGGAGAGGTGTATCGAT
CGCTAAGGCGCAACACAACGTCTCGTACGGTTGTAGTCGGATCGTCAACGTTTC
TCCCAAATCGTAGTTATCACAACTCACCGAAAGATCGGGCCAAACAACTGCCTT
GAGTGTCGAGAGGAACTCAGGGTTCATCAGGTGGTATCAGAGCTTTCGTTGCTC
GGTGAGTTTTTATCTTCCTATAACCAGAAAATAGCCATACAAAAAAAATTCGTAT
CATTTACCTAGCCATATTGTGCCATTGCATTGTTCTTTTCAGTTTTTGCTTTGTT
GAATTTTGTGTTGCATCGTCACGTCGTGTTGCTGGTCTTAGCGTCTAGTTCGTTT
AGAGTTTCGAGTTCTGGTCACGTTTTGTACCACAGACGTCGCTGCCACCGTGTC
TCTGTTGTTGTCGGGACCTACGAAAACGGAATCAGCTCGCTGCCACGGTTCTAT
TTTTGTAGTTTTCAGAAGTTTCTGTCGGTTAGTTTTGTAGTTCTTGAGTTGCATC
TAGAGTTTGCCGATCCTTGTGTGGGTTTGTTTTGGCCATGCCTCTGTCGTGCAC
GAGAGGAGGAAGGTGAATTACCATATCTGATTTTGGAAGTAGTCAAGTTTGTTT
TGGAAAGATATAGTAGATTGGATTGGTTAAAAATCAGTTTCCTTTTTATCCCATA
CACCAAATTCGGCTGCCATCCACTCCACCTCCTGGCCGAGTCCCACTCGCCCCT
TTGGCCGAGTCCGACTCCCTCCCTCTCTTCCACGATTCCGAGTTGTGTCAAACA
CCTTGGCGAATTTTTGTTTGGTGTCGGTTTCGAGATCTGTTTGGAAAAACGGAA
CCGGCATAAATTCAGCATTCCATTTTTTGTATAATTTTAATTTTGGGTTTAGACT
TTTACATTTGAGTCCCTGTAAATTTTATATTTATGTTTGAGTCCCTGTAATCTTA
CATTAAGGTCCTTGAGTCAGTTTTTGTTTAAAAAAATCAAGAAAAAAAAAGTGAG
AAAAAAAAAGGCAGAAAGGTGCAAAAAAAAAAGAGAAAAAAAAAGCCGCACAAA
AAAAACAGAAAAGAAAAGGAAAGAAAAAAAAAAAGAAAAGAAGAAAGAAGAAA
GAAAGAAAAGAGAAAATACTGTTATGTTTGAGCTGAACTTCATATATCAGACTT
GTGCATGAGTTGTTCCTAGTGCTATCTTGTGGTATCGTTTGTGTCTAGGCTCGC
GTCTCTAGTACGTTCTAGCCTAGGACCAGCACGGTACTTGACTTTGAACAATTA
TTCAACTTTGCATTATCTGATTTGAGCATTTGCTATTCCTTTGCTACATATTTAA
GCCTACCCAGAGCTCCACATATTTGATTACAGCCGTACCGCAAGTGTTTGCCAA
GGCATCGATACATCCAACCTTCATTGGGGTGCTTGGTTGAGTCGGTGTATGTCA
CCATTCCACTTGCATTGGTAAGATCTTGTAAGAGCTTGGTTAAAAGCTTGAGTG
TGTGCGATTTTTTGACCTGCCACTACCTAGTAGTTAATAGGAACGCGCATATTTT
TGTGTATGTTTCCTGTTTTCTACTAACAATGGCAGGGATACGCAAGATAATTGG
GGATAGCTGTGCTCAACATCGACATCTTCGTCGAGACATGAGGAGGGATCAACA
TGACCATTATGAGGTAAGTGATGATGTTCTAGGTAAGATCAAATCTGCACTGCC
TTATTTCGAGGGAAACTATGATCCTCGTGCTTACATTAATTGGGAGCTAGCGGT
TGATAGTGAATTTCAAAAGCATGTCTTGTCGGAGAAACAAAAGGTTATGTGTGC
CTCTAGTGTTTTAATTAAACATGCTTCTAATGATTGGAAACATCTTTGTAGGCAT
AACAAAATACCACAATCTTGGAAAGACCTGAAACGATATTTCAGAGATGTTTAT
GTTCCCATGTATTATGCTGATATTCTGCTCAACAAACTGCAATGTTTAAAACAAG
ATACCAAAAGTGTTACTTCATACTATCATGATATGCATGCTTGTTTATTACGTTG
TGGCTTAGATGAATGTGAAGAAGCTACAGAATTGAGGTTTTTACGTGGACTTAA
CAAAGAAATTCAGGACATGCTTGCTTGTGAAAAGTATAGATCTCTTTCTCATTTG
TTACAACTTGCTTGCAATGCTGAAAGTAAAATAGAGGAGGATATGAAAAAGAAA
CACGCTATGTCTTTGCCTCCAATTACTAACTATTTGCAGGAAGTGCGTAATCAT
GAAAAGGAGGAGAGAGACATGAAAGAGCCACCAATTCCATTGTTCACACTCAA
GTTCGAGACACCTCCATCATCTAAAGAGGACATCAAAGGTAAAGTAAATGGTAC
TGAAATTAATCAAGGTGAGTGCATTGTTAACGAAGTAAATTTGTTCACTTTTCAT
GCAAAAGTAGAGCAACCATTAGTGGAACCAAATGCTGGAATTC

SEQUENCE LISTING

(1) GENERAL INFORMATION:

(i) APPLICANT: Jiang, Jiming
Dong, Fenggao

(ii) TITLE OF INVENTION: DNA Sequences Specific to Rice Centromeres

(iii) NUMBER OF SEQUENCES: 8

(iv) CORRESPONDENCE ADDRESS:
(A) ADDRESSEE: Rockey, Milnamow & Katz, Ltd.
(B) STREET: 180 N. Stetson Avenue, 2 Prudential Plaza, Suite 4700
(C) CITY: Chicago
(D) STATE: Illinois
(E) COUNTRY: U.S.A.
(F) ZIP: 60601

(v) COMPUTER READABLE FORM:
(A) MEDIUM TYPE: Floppy disk
(B) COMPUTER: IBM PC compatible
(C) OPERATING SYSTEM: PC-DOS/MS-DOS
(D) SOFTWARE: PatentIn Release #1.0, Version #1.30

(vi) CURRENT APPLICATION DATA:
(A) APPLICATION NUMBER: US
(B) FILING DATE:
(C) CLASSIFICATION:

(viii) ATTORNEY/AGENT INFORMATION:
(A) NAME: Mueller, Lisa V.
(B) REGISTRATION NUMBER: 38,978
(C) REFERENCE/DOCKET NUMBER: WARF

(ix) TELECOMMUNICATION INFORMATION:
(A) TELEPHONE: 312-616-5400
(B) TELEFAX: 312-616-5460

(2) INFORMATION FOR SEQ ID NO:1:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1478 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(vi) ORIGINAL SOURCE:

(A) ORGANISM: Oryza sativa subsp. indica

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:1:

TCATCAATGA TGGGTTTGTT TTCAGAGCTA ACAAGCTATG CATTCCAGCT AGCTCCGTTC 60

GCTTGTTGTT GTTGCAGGAA GCGCATGGAG GCGATTTGAT GGGACATTTT GGTGCCAAGA 120

AGACACATGA CATCCTTGCT AGTCATTTCT TTTGGCCACA GATGCGAAGA GATGTTGGCA 180

GGTTCGTTGC TCGCTACGCT ACATGTCAAA AGGCTAAGTC ACGCTTACAT CCACATGGTT 240

TGTATATGCC TCTTCCTGTT CCTACTGTTC CTTGGGAAGA TATTTCAATG GATTTTGTGT 300

TAGGATTGCC TAGGACCAAG AGGGGGCGTG ATAGCATTTT TGTGGTTGTG GATCGATTTT 360

CTAAAATGGC ACATTTCATA CCATGTCATA AAACTGATGA TGCTTCTCAT ATCGCTGATT 420

TGTTCTTTCG AGAAATTGTT CGCTTGCATG GTGTGCCAAA CACAATTGTT TCTGATCGTG 480

ACACAAAATT TCTTAGCCAT TTTTGGAGAA CTTTGTGGGC TAAATTGGGG ACTAAACTTT 540

TGTTTTCTAC TACTTGTCAT CCCCAAACTG ATGGACAAAC TGAAGTGGTG AATAGAACCT 600

TGTCTACTAT GCTTAGGGCT GTTTGAAGAA AAATATCAAG ATGTGGGAAG AATGCTTGCC 660

TCATATTGAA TTTGCTTATA ATCGTTCCTT GCATTCTACT ACAAAAAATG TGCCCATTTC 720

AGATTGTGTA TGGTTTGTTG CCTCGTGCTC CAATTGATTT GATGCCTTTA CCATCTTCTG 780

AGAAACTGAA TTTTGATGCG AAGCAACGTG CTGAGTTGAT GTTAAAACTG CATGAGACAA 840

CTAAAGAAAA CATAGAGCGC ATGAATGCTA AGTATAAGTT TGCTGGTGAC AAAGGTAGAA 900

GGGAATTGAA TTTTGAACCT GGAGATTTGG TTTGGTTGCA TTTGCGAAAA GAACGATTTC 960

CTGATTTGAG AAAGTCTAAA TAGATGCCTA GAGCTGATGG ACCATTTAAA GTGTTAGCAA 1020

AGATTAATGA GAATGCATAT AAGATTGATT TGCTTGCAGA TTTCGGGGTT AGTCCCACAT 1080

TTAACATTGC AGATTTGAAG CCGTATATTG GGAGAAAAAG ATGAGCTTGA GTCGAGGATG 1140

ACTCAAATGC AAGATGGGGA GGATGATGAG GACATCAACA CCATCGATAC ATCCACGTCC 1200

CCCCATATAC AGCATGATGG TCCTATTACC CGCGCTTGTG CACGTCAACT AAATTATCAG 1260

GTGATTCTTT CTTGAGTTCA AATTTCCTCG TCTTTATACC TCGGAGACGC GTGCACTCGT 1320

GTTTTACTCC AGGAACGTAT GGAGAGGATC AAAGGGAAGA GGATTCGCGC GGGGTGGATT 1380

CGGACTGCAG GGCAGCGCCA ACTTCTGACG GCCGCCACGA CTTCATGCAA ACTCCGATTT 1440

GGGCGTGCAA GTACTTCATG GAAAGCTTAT CAAGTCTA 1478

(2) INFORMATION FOR SEQ ID NO:2:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 639 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(vi) ORIGINAL SOURCE:
(A) ORGANISM: Oryza sativa subsp. indica

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:2:

GATCTTTTCT ACTGGAATCA AAATGTTCAA AAAATGCCAA AACATGATTT TTGGACATAT 60

TGGAGTGTAT TGGGTGCATT CATGGCAAAA ACTCACTCCG TGATTCGCGC GGCGAACTTT 120

TGTCAATTAA TGCCAAIATT GGGACACGAG GGTGCGATGT TTTCTACTGG AATCAAAAAG 180

TTCAAAAAGA GCCAAAACAT GATTTTTGGA CATATTGGAG TGTATTGGGG TGCGTTCGTG 240

GCAAAACTCA CTTCGTGATT CGCGCGGCGA ACTTTTGTCA TTTAATGCCA ATATGTGCAT 300

ACACGAGAGA GTGCGATGTT ATTCTACTGG AATCAAAAAG TTCAAAAAGA GCCAAAACAT 360

GATTTTTGGA CATATTGGAG TGTATTGGGT GCGTTCGTGG CAAAACTCAC TTCGTGATTC 420

GCGCGGCGAA CTTTTGTCAA TTAATTCCAA TATGTGCATA TTTTGGCCCA AAGTGTTGCG 480

ATGTTATTAA TCGGAATGAA AAAGTTCAAA AGGCACCAAA ACATGATTTT TTGACATATT 540

GGAGTGTATT GGGTGCGTTC GTGGGAAAAA CTCACTTCGT GATTCGCGCG GCGAACTTTT 600

GTCATTTAAT GCCAATATTG GCACAGCGAC GGGTGCGAT 639

(2) INFORMATION FOR SEQ ID NO:3:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 827 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(vi) ORIGINAL SOURCE:
(A) ORGANISM: Oryza sativa subsp. indica

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:3:

AAGCTTCATA CTTTGTCCGT TTGAAATAAG ATTTTCTGTG GTTGGTTCTA GTGTGCTCCT 60

GGTTGTGAAA AAAAATATCA AAAAAATAAT AAAAAAATAA AAAAAAAGTT GTGCCTCTCT 120

CCTCAGCACT AGAGAGAGCA AGCAAAAAAA AAAAAAGAGG ATCCAGCGAG CACACCACAC 180

CATTCCCCCT TTGTCACCAC CTTCCACCAC CGGTTCTTGT TTTCGTTGCT GTTTGGGTTT 240

GTGTTTTGTG GCGTCCTTTG TGGTTAGACT CGTCGTCTCT ACTTCGACTA GCCTAGGACC 300

AGCTACTGTA CACTATACCC TCTGAGCGAT TATTCAATTT GCTTTGGCTA ACGTGGTTTC 360

TAATTCTTCT CTTGAGCAAT GCTTTCAGCC CACCAACAGC TCCACCACAA ACCCGACTAC 420

AGCTTGACAG GTCTTGTTGC TACAGCAACG ATACACCTCG TTCCACGGTA GCTAACTAGT 480

GGTGTTGCTG TCCCCTACCT GTGGGCAAGG TAAGAACTGG TAAGAGCTTG TGAGACAAGC 540

TGCGAGTGAA GTGAGAGTGA GCGTCTTGCA AAAGCTACAT CCCAATAGTT GTGTAGGGCT 600

TACATTACAT CCTCCATTGT TTTGTTGTCC TTCATTTCTA AACCATGGCA GGTCCAGAGG 660

ATAAAGATGA GTAGGGTGCA TCTCATTCCC CACGCACCAA AGGCATCATC CAATACTTTA 720

CAAGGCAAGT GAAGCAGCAC ACAGAAGGAC TTGATACAGA TTTGCAGGTG ACAAATGAGA 780

AGATTGGACA ATTGGAGTCC ACGCAGATCT CCACCAACAC TAAGCTT 827

(2) INFORMATION FOR SEQ ID NO:4:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1201 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(vi) ORIGINAL SOURCE:
(A) ORGANISM: Oryza sativa subsp. indica

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:4:

AAGCTTCTCG CCACAAAAAC AGAATTTTGC AAGAGGAATT TGGCGATTTT AAATTTGCCG 60

AACCCAGCAA AGCTGGGGAG AAACGGATGT AACTTTTTTT GTAGATGTCC GTGGATATAT 120

TATAAGCCTA ATTTCGAGTC CGTATGAAAA AGATATGACT GTTTTAAGGA AGGTCACTCG 180

AATCAGGGTT CTATTTCCTT TCTAGTGGCG GCAGGATACA AGGCAACTCA CAAAAGCTAT 240

CGGCAACAGA TCAAACACAA ATAGAAAGTG ATTTGGTGGG TGACTGCGAC GGTGGTGACG 300

ATGGTGATGA CGGCAAAATT GATGAAGTGG CGATGTGCGG TGGTGGCAAG AACTCGAAAC 360

TCTAATCGCC TAACACTAAG ACCAGCAAAT TAATAAAACC GATGCAATCG AAAACTCAAC 420

AAGCCTTAAC TAAGCAGTAC TAATGAAGAG GCACAATGGT ATATATGATC ACCGAAAAAC 480

TAATCTACTA TTTTTTTGTT GTTGTCTTTT CTGGACTATA GGAAAAAAAA ATAATGACGA 540

AGGGAAGGGT AAATTCTCTC ACCGATGAAC CACGCTCTGC TACCACCTGA TGCGAACCCG 600

CTAGGGTTTG TGGCCCGATC TTTCGATGAG AGGTAAGGGA TAACTCGATT AGATGAAGTC 660

GACGTTCACG GCCCCACTAC AACTGTCCAA AGACGCTGTG CCTTACCAAC CGATACACCG 720

TCTCCAATGG TCGCCGCGAA CTTGTGGTGC GCGTCAACCC GGCCACGAGG GCACTCGTCC 780

TGCAAGCAAT CGAAGAACAA GCAAGAACAA GTAGGACAAG CACTGAAATT GCTAGATAAA 840

GATGAAAGTT TCACTATCAA ACACAATATG GTGGGGTTCC GATTACAGGC AGACGCGGCG 900

GTTTAGCCAA CACGCGCGCT GCGAGCCGGT AGCAAGAAGC TATCTTCTAT CATCAAAACC 960

CGCCTGTTTT TGGCGGCGAC TAGAGGTATA ACACAAGAGG GGAAGAACGA CCATAGGGTC 1020

GTGCTCCAAC CCTAGGACGC GCCCCTAATG GGCCCAACAT GGATACACAG CCCATTGGGC 1080

CAAAAGAGGT GACGCAGCAC CATGGACAGA AAATAGTCGG GAGTAAAATG ACAATTGCGG 1140

CCGCTCCAGA ACAGATATGG ACATGTGGCT GGATCCACTT GAAAGTAGAC TTGATAAGCT 1200

T 1201

(2) INFORMATION FOR SEQ ID NO:5:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1341 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(vi) ORIGINAL SOURCE:
(A) ORGANISM: Oryza sativa subsp. indica

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:5:

AAGCTTGTCG CCCGCGGCCA CTTTAGNCTT GCTGCTCTTG TGACGCATTT TGATGCTCTC 60

AATGCTGGAG GCAATGGTGG GGGTAATGAT GATGACATTG ACGGAGAGTA CGTCGAGGAT 120

AATTTGGAGG ATGAATATAT TGCTGACACT GAACAAGATG ATCGAGATGC TCGAGATCGT 180

CGACGGCTAC ACAACAATCG ACGAGGTATG GGTGGTCGCC GCCGACGCGA GGTACGCAAC 240

AATGATGATG CTTTTTCTAA AATTAAATTT AAGATTCCTC CTTTTGATGG AAAATATGAT 300

CCTGATGCGT ACCTCTCTTG GGAGATTGCT GTTGACCAAA AATTTGCATG CCATGAGTTT 360

CCTGAGAGCA CTAGAGTTAG AGCTGCTACG AGTGAGTTCA CCGATTTTGC TTTGGTTTGG 420

TGGATAGAAC ATGGCAAGAA AAATCCTAAT AACATGCCAC AAACTTGGGA TGCTTTGAAA 480

CGGGTCATGC GGGCTAGATT TGTTCCTTCT TACTATGCAC GCGACTTGCT GAATAGGTTG 540

CAACAATTGA GACAGGGTGC GAAAAGTGTA GAGGAATATT ATCAGGAGTT ACAAATGGGC 600

TTGCTTCGTT GTAATTTAGA GGAAACTGAG GACGCTGCCA TGGCTAGATT TTTGGGTGGG 660

TTAAACCGCG AGATTTATGA CATCGTAGAC TATAAAGATT ACGCTAATAT GACCCGATTG 720

TTTCATCTTG CTTGTAAGGC TGAAAGGGAA GTGCAAGGAC GACGTGCTAG TGCCAAGGCT 780

AATTTTTCTG CAGGTAAAAC TTCATCATGG CAGACACGCA CCACTCCTCC GGCCGGCCGT 840

ACTGCTTCTC CATCTTCCAC ACCCACAACC AGTCGAGCAG CACCTCCTCC ATCTAGTGAC 900

AAGTCAGCGA CAAAGGCTGC TCAGCCAGCA CCGAGTGCTT CTTCAATGGC ATCCACAGGC 960

CGAATGAGAG ATGTTCAGTG CCACCGTTGC AAGGGCTTTG GGCATGTGCA GCGTGACTGC 1020

CCTAGCAAGC GAGTTTTGGT AGTCAAAAAC GATGGTGAGT ACTCCTCTGC TAGTGATTTC 1080

GATGATGATA CACTTGCTTT GCTTGCGGCT GACCATGCAG ATAATGAGCC ACCGGAAGAG 1140

CACATTGGGG CTGCATTTGC GGATCACTAT GAGAGCCTCA TTGTGCAGCG TGTCCTTAGC 1200

GCACAAATGG AGAAGGCGGA GCAAAATCAG CGACACACGT TGTTCCAAAC AAAGTGTGTC 1260

GTCAAAGAGC GTTGTTGCCG CATGATCATT GATGGAGGTA GCTGCAACAA CTTGGCTAGC 1320

AGCGAGATGG TGGAGAAGCT T 1341

(2) INFORMATION FOR SEQ ID NO:6:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 701 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(vi) ORIGINAL SOURCE:
(A) ORGANISM: Oryza sativa subsp. indica

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:6:

GAATTCCTTT GTCACAAGTT GATTTACTCG CTGTTCCTTG TGATAAAGAA GAGTTGTGTG 60

ATAATGCTTC ACTTAAATCC ATGCCACAAC TAGTGAATGA ACATGCTATT CCTAGTGTAT 120

CTTTCTGTGC TGATTTTAAG CATGTTGTTC ACATTGCTAA TAGAGTAGAG GAACGTGAAT 180

TGACATCTTC TTTAAATACT TTGGGCTATG TTCAGTTTGC TGATTTTCAT GAGCTCGATA 240

ATTTGAAGGA GAAATTATTT GCTAAGTCTG ATTTGCCATG TCCAAGTAAC GCTATTTTTC 300

ATCTCTTTGG TGAATATAAT GATAGAGGAA TATATTTGGT GCATAGAGTT TACATCTGTT 360

CAGATTTAGA ACCTCCTGTA CATGTGGATA AAACATGCAA GCTAGAGAGA AATGTTATTG 420

CTAACAAAAT TGTCTCGAGT TTGTCTTGTT TTGATTGGAC GAAACAGGTT GTTGTAGAAT 480

GGTACTACGC AATAGAGCAC CACTATGGAG ATAATACCGA GGACGGTTTT CCATTGAAGA 540

TACGGGGAGG TATGATGTGA CCATGGCTAC ACACGGATGC AACTCATGTG ACTCACATGC 600

ATAGGATGAA CAGAGAAGAT ATCGAGATCA CATCGTCCAA GTGCTGGAAT CCGATTCGGC 660

CACACTACTA CTACTGCTGA CTTCAAACGG CCGCCGAATT C 701

(2) INFORMATION FOR SEQ ID NO:7:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 2863 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(vi) ORIGINAL SOURCE:
(A) ORGANISM: Oryza sativa subsp. indica

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:7:

GAATTCCATC TCGTGTGGCC GGAATCACAG GCACAAGGTG CTGCGTCACC TATAATGGGC 60

CTATGGGCTT GTAACTTCAT TTGGGACCCC GGCCCAGGTG GGGCCCATGT GGGGTGCGCC 120

CCAACCAGGT GGAGCACGAC CCTAGGCACC CCTAGGTCGT CCTCCCACCC CTATATATAG 180

CTAGGTACCC CTTCAGGGTT TCTTGGGTTT TGATAGATTA AAGTTTAGCC ATTGCTACTT 240

GCTTGCAGCG CGCGTGTCGC CTAGACCGTC CGTCTGCTTG TTCTTCGGAA CCCCAACTTA 300

TCATTTGTAT TAAATTCCTA TTTGCAATAT CAGATTGCTT TTATCTTGTT CTTGCTTGTT 360

TCTTCGATTT GCTTGCAGGA ATAGGGTTGA TCTGCACCGG TAAGATCAAC AACCCACGGA 420

GAGGTGTATC GATCGCTAAG GCGCAACACA ACGTCTCGTA CGGTTGTAGT CGGATCGTCA 480

ACGTTTCTCC CAAATCGTAG TTATCACAAC TCACCGAAAG ATCGGGCCAA ACAACTGCCT 540

TGAGTGTCGA GAGGAACTCA GGGTTCATCA GGTGGTATCA GAGCTTTCGT TGCTCGGTGA 600

GTTTTTATCT TCCTATAACC AGAAAATAGC CATACAAAAA AAATTCGTAT CATTTACCTA 660

GCCATATTGT GCCATTGCAT TGTTCTTTTC AGTTTTTGCT TTGTTGAATT TTGTGTTGCA 720

TCGTCACGTC GTGTTGCTGG TCTTAGCGTC TAGTTCGTTT AGAGTTTCGA GTTCTGGTCA 780

CGTTTTGTAC CACAGACGTC GCTGCCACCG TGTCTCTGTT GTTGTCGGGA CCTACGAAAA 840

CGGAATCAGC TCGCTGCCAC GGTTCTATTT TTGTAGTTTT CAGAAGTTTC TGTCGGTTAG 900

TTTTGTAGTT CTTGAGTTGC ATCTAGAGTT TGCCGATCCT TGTGTGGGTT TGTTTTGGCC 960

ATGCCTCTGT CGTGCACGAG AGGAGGAAGG TGAATTACCA TATCTGATTT TGGAAGTAGT 1020

CAAGTTTGTT TTGGAAAGAT ATAGTAGATT GGATTGGTTA AAAATCAGTT TCCTTTTTAT 1080

CCCATACACC AAATTCGGCT GCCATCCACT CCACCTCCTG GCCGAGTCCC ACTCGCCCCT 1140

TTGGCCGAGT CCGACTCCCT CCCTCTCTTC CACGATTCCG AGTTGTGTCA AACACCTTGG 1200

CGAATTTTTG TTTGGTGTCG GTTTCGAGAT CTGTTTGGAA AAACGGAACC GGCATAAATT 1260

CAGCATTCCA TTTTTTGTAT AATTTTAATT TTGGGTTTAG ACTTTTACAT TTGAGTCCCT 1320

GTAAATTTTA TATTTATGTT TGAGTCCCTG TAATCTTACA TTAAGGTCCT TGAGTCAGTT 1380

TTTGTTTAAA AAAATCAAGA AAAAAAAGTG AGAAAAAAAA AGGCAGAAAG GTGCAAAAAA 1440

AAAGAGAAAA AAAAAGCCGC ACAAAAAAAA CAGAAAAGAA AAGGAAAGAA AAAAAAAAAG 1500

AAAAGAAGAA AGAAGAAAGA AAGAAAAGAG AAAATACTGT TATGTTTGAG CTGAACTTCA 1560

TATATCAGAG TTGTGCATGA GTTGTTCCTA GTGCTATCTT GTGGTATCGT TTGTGTCTAG 1620

GCTCGCGTCT CTAGTACGTT CTAGCCTAGG ACCAGCACGG TACTTGACTT TGAACAATTA 1680

TTCAACTTTG CATTATCTGA TTTGAGCATT TGCTATTCCT TTGCTACATA TTTAAGCCTA 1740

CCCAGAGCTC CACATATTTG ATTACAGCCG TACCGCAAGT GTTTGCCAAG GCATCGATAC 1800

ATCCAACCTT CATTGGGGTG CTTGGTTGAG TCGGTGTATG TCACCATTCC ACTTGCATTG 1860

GTAAGATCTT GTAAGAGCTT GGTTAAAAGC TTGAGTGTGT GCGATTTTTT GACCTGCCAC 1920

TACCTAGTAG TTAATAGGAA CGCGCATATT TTTGTGTATG TTTCCTGTTT TCTACTAACA 1980

ATGGCAGGGA TACGCAAGAT AATTGGGGAT AGCTGTGCTC AACATCGACA TCTTCGTCGA 2040

GACATGAGGA GGGATCAACA TGACCATTAT GAGGTAAGTG ATGATGTTCT AGGTAAGATC 2100

AAATCTGCAC TGCCTTATTT CGAGGGAAAC TATGATCCTC GTGCTTACAT TAATTGGGAG 2160

CTAGCGGTTG ATAGTGAATT TCAAAAGCAT GTCTTGTCGG AGAAACAAAA GGTTATGTGT 2220

GCCTCTAGTG TTTTAATTAA ACATGCTTCT AATGATTGGA AACATCTTTG TAGGCATAAC 2280

AAAATACCAC AATCTTGGAA AGACCTGAAA CGATATTTCA GAGATGTTTA TGTTCCCATG 2340

TATTATGCTG ATATTCTGCT CAACAAACTG CAATGTTTAA AACAAGATAC CAAAAGTGTT 2400

ACTTCATACT ATCATGATAT GCATGCTTGT TTATTACGTT GTGGCTTAGA TGAATGTGAA 2460

GAAGCTACAG AATTGAGGTT TTTACGTGGA CTTAACAAAG AAATTCAGGA CATGCTTGCT 2520

TGTGAAAAGT ATAGATCTCT TTCTCATTTG TTACAACTTG CTTGCAATGC TGAAAGTAAA 2580

ATAGAGGAGG ATATGAAAAA GAAACACGCT ATGTCTTTGC CTCCAATTAC TAACTATTTG 2640

CAGGAAGTGC GTAATCATGA AAAGGAGGAG AGAGACATGA AAGAGCCACC AATTCCATTG 2700

TTCACACTCA AGTTCGAGAC ACCTCCATCA TCTAAAGAGG ACATCAAAGG TAAAGTAAAT 2760

GGTACTGAAA TTAATCAAGG TGAGTGCATT GTTAACGAAG TAAATTTGTT CACTTTTCAT 2820

GCAAAAGTAG AGCAACCATT AGTGGAACCA AATGCTGGAA TTC 2863

(2) INFORMATION FOR SEQ ID NO:8:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 745 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(vi) ORIGINAL SOURCE:
(A) ORGANISM: Sorghum bicolor

(vii) IMMEDIATE SOURCE:
(B) CLONE: pSau3A9

(x) PUBLICATION INFORMATION:

(A) AUTHORS: Jiang, Jiming
Nasuda, Shuhei
Dong, Fenggao
Scherrer, Christopher W.
Woo, Sung-Sick
Wing, Rod A.
Gill, Bikram S.
Ward, David C.
(B) TITLE: A Conserved Repetitive DNA Element Located in Centromeres of Cereal Chromosomes
(C) JOURNAL: Proc. Natl. Acad. Sci. U.S.A.
(D) VOLUME: 93
(F) PAGES: 14210-14213
(G) DATE: November-1996

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:8:

GATCTTTGGA TTGGAAACAG TTAAACGGCA ATATGTGCAT GATGATGATT TTAAAGATGT 60

GTTTTTGCAT TGTAAGGATG GGAAGGCATG GAATAAATTT GTTGTAAATG ATGGTTTTGT 120

GTTTAGAGCT AATAAGCTAT GCATTCCAGC TAGCTCTGTT CGTTTGTTGT TGCTACAGGA 180

AGCACATGGA GGTGGTTTGA TGGGACATTT TGGGGCAAAG AAGACGGAGG ACATACTGGC 240

TGGTCATTTC TTTTGGCCAA AGATGAGGAG AGATGTGGAG AGATTTATTG CTCGCTGCAC 300

GACATGTCAA AAGGCCAAGT CACGCTTAAA TCCACACGAT TTGAAGCCAT ATTTGGGTGA 360

GGGAGATGAG CTTGAGTCGG GGACGACTCA AATGCAAGAA GGGGAGGATG ATGAGGACAT 420

CAGCACCATC TATACATCCA CACCTACACC CACACCATCG CCAACACCAC TTGGCCCTCT 480

TACTCGTGCC AGTGCCCGTC AACTGAACCA TCAAGTAAGT TTATTCTTAA ACTCTTGTCC 540

ATCATATTTA GACAATGGAG ACACGTGCAC TCTTGTTTTG CTTAGGAATG ATGGAGAGGA 600

CCAGAAGCAT AGGGGATTGG TGTAGGCTGG ATTTGGACAG CAAGACAGCA CCAACTTACA 660

ACAACCGCCA TGACTTCATA CAGAGTCCAT TTTAAGCATG CAAGCACTTG ATGGAAAACT 720

CGTCAAGTAT ATTTTTAGAT GGATC 745

# INTERNATIONAL SEARCH REPORT

Inter nal Application No
PCT/US 00/17535

A. CLASSIFICATION OF SUBJECT MATTER
IPC 7 C12N15/82 C12N5/10 C12Q1/68 A01H5/00

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)
IPC 7 C12N C12Q A01H

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practical, search terms used)
EPO-Internal, STRAND, WPI Data, PAJ, BIOSIS

C. DOCUMENTS CONSIDERED TO BE RELEVANT

| Category ° | Citation of document, with indication, where appropriate, of the relevant passages | Relevant to claim No. |
|---|---|---|
| X | DONG FENGGAO ET AL: "Rice (Oryza sativa) centromeric regions consist of complex DNA." PROCEEDINGS OF THE NATIONAL ACADEMY OF SCIENCES OF THE UNITED STATES, vol. 95, no. 14, 7 July 1998 (1998-07-07), pages 8135-8140, XP002153043 July 7, 1998 ISSN: 0027-8424 cited in the application | 1,22 |
| Y | the whole document | 2,8-13, 19,21 |
| | --- -/-- | |

[X] Further documents are listed in the continuation of box C.

[X] Patent family members are listed in annex.

° Special categories of cited documents:

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier document but published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art.

"&" document member of the same patent family

| Date of the actual completion of the international search | Date of mailing of the international search report |
|---|---|
| 28 November 2000 | 11/12/2000 |

| Name and mailing address of the ISA | Authorized officer |
|---|---|
| European Patent Office, P.B. 5818 Patentlaan 2<br>NL - 2280 HV Rijswijk<br>Tel. (+31-70) 340-2040, Tx. 31 651 epo nl,<br>Fax (+31-70) 340-3016 | Holtorf, S |

Form PCT/ISA/210 (second sheet) (July 1992)

# INTERNATIONAL SEARCH REPORT

Inter nal Application No
PCT/US 00/17535

C.(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT

| Category ° | Citation of document, with indication,where appropriate, of the relevant passages | Relevant to claim No. |
|---|---|---|
| X | DATABASE EMBL SEQUENCE DATABASE 'Online! 3 December 1998 (1998-12-03) WING, R.A. AND DEAN, R.A.: "a BAC end sequencing framework to sequence the rice genome" XP002153047 accession no. AQ291790 --- | 1 |
| Y | NONOMURA K -I ET AL: "Organization of the 1.9-kb repeat unit REC1 in the centromeric region of rice chromosomes." MOLECULAR AND GENERAL GENETICS, vol. 261, no. 1, February 1999 (1999-02), pages 1-10, XP002153044 ISSN: 0026-8925 page 2, left-hand column --- | 2,8-13, 19,21 |
| Y | JIANG JIMING ET AL: "A conserved repetitive DNA element located in the centromeres of cereal chromosomes." PROCEEDINGS OF THE NATIONAL ACADEMY OF SCIENCES OF THE UNITED STATES, vol. 93, no. 24, 1996, pages 14210-14213, XP002153045 1996 ISSN: 0027-8424 cited in the application abstract, Fig. 4; page 14213, right column, last paragraph --- | 2,8-13, 19,21 |
| Y | DATABASE EMBL SEQUENCE DATABASE 'Online! 6 January 1999 (1999-01-06) DONG, F., ET AL.: "Rice (Oryza sativa) centromeric regions consist of complex DNA" XP002153048 accession no.AF078903 --- | 13,19,21 |
| Y | DATABASE EMBL SEQUENCE DATABASE 'Online! 20 July 1998 (1998-07-20) DONG, F., ET AL.: "Rice (Oryza sativa) centromeric regions consist of complex DNA." XP002153049 accessio no. AF058902 --- | 13,19,21 |
| Y | DATABASE EMBL SEQUENCE DATABASE 'Online! 20 July 1998 (1998-07-20) DONG, F., ET AL.: "Rice (Oryza sativa) centromeric regions consist of complex DNA." XP002153050 accession no. AF058903 --- -/-- | 13,19,21 |

3

# INTERNATIONAL SEARCH REPORT

Inter nal Application No
PCT/US 00/17535

C.(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT

| Category * | Citation of document, with indication,where appropriate, of the relevant passages | Relevant to claim No. |
|---|---|---|
| Y | DATABASE EMBL SEQUENCE DATABASE 'Online! 20 July 1998 (1998-07-20) DONG, F., ET AL.: "Rice (Oryza sativa) centromeric regions consist of complex DNA." XP002153051 accession no. AF058904 --- | 13,19,21 |
| Y | DATABASE EMBL SEQUENCE DATABASE 'Online! 20 July 1998 (1998-07-20) DONG, F., ET AL.: "Rice (Oryza sativa) centromeric regions consist of complex DNA." XP002153052 accession no. AF058905 --- | 13,19,21 |
| Y | DATABASE EMBL SEQUENCE DATABASE 'Online! 20 July 1998 (1998-07-20) DONG, F., ET AL.: "Rice (Oryza sativa) centromeric regions consist of complex DNA." XP002153053 accession no. AF058906 --- | 13,19,21 |
| A | WO 98 55637 A (ARCH DEV CORP ;PREUSS DAPHNE (US); COPENHAVER GREGORY (US)) 10 December 1998 (1998-12-10) claim 36 --- | 1-22 |
| A | WO 89 09219 A (GEN HOSPITAL CORP) 5 October 1989 (1989-10-05) cited in the application the whole document --- | 6,16,17 |
| A | ARAGON-ALCAIDE LUIS ET AL: "A cereal centromeric sequence." CHROMOSOMA (BERLIN), vol. 105, no. 5, 1996, pages 261-268, XP000965076 ISSN: 0009-5915 the whole document --- | |
| P,X | DATABASE EMBL SEQUENCE DATABASE 'Online! 31 January 2000 (2000-01-31) SASAKI, T. , ET AL.: "Oryza sativa nipponbare (GA3) genomic DNA, chromosome 1, PAC clone: P0499C11" XP002153054 accession no AP001080 ----- | 1 |

3

Form PCT/ISA/210 (continuation of second sheet) (July 1992)

# INTERNATIONAL SEARCH REPORT

Information on patent family members

Inter nal Application No
PCT/US 00/17535

| Patent document cited in search report | Publication date | Patent family member(s) | Publication date |
|---|---|---|---|
| WO 9855637 A | 10-12-1998 | AU 7608698 A | 21-12-1998 |
| | | EP 0986648 A | 22-03-2000 |
| WO 8909219 A | 05-10-1989 | AU 622900 B | 30-04-1992 |
| | | AU 3346489 A | 16-10-1989 |
| | | EP 0338266 A | 25-10-1989 |
| | | HU 55052 A | 29-04-1991 |
| | | IL 89453 A | 29-06-1995 |
| | | JP 3503599 T | 15-08-1991 |
| | | US 5270201 A | 14-12-1993 |
| | | ZA 8901809 A | 29-11-1989 |